

AFGL-TR-81-0038

LEVEL

12

AD A104178

NUMERICAL METHODS FOR HARMONIC ANALYSIS
ON THE SPHERE

Oscar L. Colombo

The Ohio State University
Research Foundation
Columbus, Ohio 43212

DTIC
S ELECTE D
SEP 14 1981
H

March 1981

Scientific Report No. 7

Approved for public release; distribution unlimited

AIR FORCE GEOPHYSICS LABORATORY
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
HANCOM AFB, MASSACHUSETTS 01731

DTIC FILE COPY

81 9 14 095

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM	
1. REPORT NUMBER AFGL/TR-81-0038	2. GOVT ACCESSION NO. AD-A104 178	3. RECIPIENT'S CATALOG NUMBER Sci - 310	
4. TITLE (and Subtitle) NUMERICAL METHODS FOR HARMONIC ANALYSIS ON THE SPHERE		5. TYPE OF REPORT & PERIOD COVERED Scientific Report No. 7	
		6. PERFORMING ORG. REPORT NUMBER Dept. of Geod. Sci. No. 310	
7. AUTHOR(s) Oscar L. Colombo		8. CONTRACT OR GRANT NUMBER(s) F19628-79-C-0027	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Geodetic Science The Ohio State University - 1958 Neil Avenue Columbus, Ohio 43210		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2309G1AW	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Geophysics Laboratory Hanscom AFB, Massachusetts 01730 Contract Monitor: Bela Szabo/LW		12. REPORT DATE March 1981	
		13. NUMBER OF PAGES 140	
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Sci - 310, Scientific-7		15. SECURITY CLASS. (of this report) Unclassified	
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE	
16. DISTRIBUTION STATEMENT (of this Report) A - Approved for public release; distribution unlimited			
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)			
18. SUPPLEMENTARY NOTES			
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) geodesy, gravity, harmonic analysis			
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This report presents some numerical methods for estimating spherical harmonic coefficients from data sampled on the sphere. The data may be given in the form of area means or of point values, and it may be free from errors or affected by measurement "noise". The case discussed to greatest length is that of complete, global data sets on regular grids (i.e., lines of latitude and longitude, the latter, at least, separated by constant interval); the case where data are sparsely and irregularly distributed is also considered in some detail. →			

DD FORM 1 JAN 73 1473 K EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

400254

The first section presents some basic properties of spherical harmonics, stressing their relationship to two-dimensional Fourier series. Algorithms for the evaluation of the harmonic coefficients by numerical quadratures are given here, and it is shown that the number of operations is the order of N^3 for equal angular grids, where N is the number of lines of latitude, or "Nyquist frequency", of the grid.

The second section introduces a quadratic measure for the error in the estimation of the coefficients by linear techniques. This is the error measure of least squares collocation, which is a method that can be used for harmonic analysis. Efficient algorithms for implementing collocation on the whole sphere are described. A formal relationship between collocation and least squares adjustment is used to obtain an alternative form of the collocation algorithm that is likely to be stable with dense data sets and, with a minor modification, can be used to implement least squares adjustment as well. The basic principle is that for regular grids the variance-covariance matrix of the data consists of Toeplitz-circulant blocks, so it can be both set up and inverted very efficiently.

Section four describes a method for computing the covariances between area means that is much more efficient than the usual approach of integrating numerically the point covariance function twice over the blocks. This method is essential for setting up the variance-covariance matrix of the data to implement collocation when the data are area means. In this and other respects the present work is complementary to Report 291, by the same author.

Finally, there is an Appendix containing the listing of subroutines that implement various algorithms, together with some explanatory notes on how to use them.

Foreword

This report was prepared by Dr. Oscar L. Colombo, Post Doctoral Researcher, Department of Geodetic Science, The Ohio State University, under Air Force Contract No. F19628-79-C-0027, The Ohio State University Research Foundation Project No. 711664, Project Supervisor Richard H. Rapp. The contract covering this research is administered by the Air Force Geophysics Laboratory, Hanscom Air Force Base, Massachusetts, with Mr. Bela Szabo, Contract Monitor.

Accession For	
NTIS GR&I	
DTIC TAB	
Unannounced	
Justification	
By	
Distribution	
Availability	
DTIC	
A	

Acknowledgements

I am indebted to Dr. R. H. Rapp for his support during the long months when this report was gestating; to K. Katsambalos and C. Jekeli for comments, particularly on Section 1; to Pamela Hager for typing the beginning of section 1, and section 3; and to Susan L. Carroll for typing all the rest, a hard task that she has done with care and good cheer.

The ideas presented in the pages that follow began to develop while I was working for the late Dr. R. S. Mather at the University of New South Wales, Australia. They continued to grow, out of discussions with him, his students, and many others, first in Australia and then in the USA. Those discussions made clear to me that there was (and still is, of course) real need to find efficient ways for handling the large amounts of data that are becoming available to Earth scientists on a global basis, chiefly through the use of space technology. To all of them, my sincere thanks for their beneficial influence.

Table of Contents

Foreword	iii
Acknowledgements	iv
1. Introduction	1
1.1 Spherical Harmonic Analysis and Synthesis: Definitions	2
1.2 Relationships between Spherical Harmonics and 2-D Fourier Series	6
1.3 Sampling Errors	9
(a) Number of equations and number of unknowns	11
(b) 100% aliasing	12
(c) Orthogonality	12
1.4 Number of Operations in Analysis and in Synthesis	13
1.5 Algorithm for the Analysis of Point Values	15
1.6 Algorithm for the Synthesis of Point Values	17
1.7 Algorithms for the Analysis and Synthesis of Area Means	18
1.8 Duality between Analysis and Synthesis	18
1.9 Usefulness of the Fast Fourier Transform Method	19
1.10 Functions Harmonic in Space and their Gradients	21
2. Error Measure and Optimal Quadrature Formulas	23
2.1 The Isotropic Covariance	23
2.2 Some Additional Notation	26
2.3 Estimation Errors, Sampling Errors, and Propagated Noise .	29
2.4 The Quadratic Error Measure.	30
(a) Propagated noise measure	30
(b) Sampling Error Measure	31

(c) Total Error Measure	31
2.5 The Meaning of the Error Measure	32
2.6 Simple Formulas for Area Means	33
2.7 Optimum de-Smoothing Factors	35
2.8 Least Squares Collocation	36
2.9 The Best Quadrature Formula for non-Uniform, Uncorrelated Noise	38
2.10 The Structure of the Covariance Matrix and its Consequences	40
2.11 Setting up and Inverting the Covariance Matrix	43
2.12 Optimal Formulas for non-Uniform, Correlated Noise	46
2.13 Least Squares Adjustment and Least Squares Collocation	47
(a) Band-limited signal	47
(b) Infinite bandwidth	50
2.14 Ridge Regression and Least Squares Collocation	51
2.15 Structure of the Normal Matrix	53
2.16 Global Adjustment and Collocation with Scattered Data	56
(a) Full region bound by lines of latitude and longitude	56
(b) Arbitrarily scattered data	57
2.17 The Error Matrix in the Band Limited Case	63
2.18 The Use of a Priori Information of the Coefficients	63
(a) Simple weighted averages	64
(b) A priori values included as data	65
(c) The method of Kaula and Rapp	65
2.19 Optimal Estimation over a Band of Spatial Frequencies . .	66

3. Numerical Examples	68
3.1 Generation and Analysis of Simulated Data	68
3.2 Agreement between the Actual and the Theoretical Measures of the Sampling Errors	71
3.3 Accuracies of Various Quadratures Formulas	71
3.4 The Analysis of a Global Data Set of $5^\circ \times 5^\circ$ Mean Anomalies ..	80
4. Covariances between Area Means	82
4.1 Derivation of an Approximate Formula for the Covariance ..	82
4.2 Choosing N_{\max}	84
4.3 Numerical Examples	85
4.4 Covariances between Mean Values and Point Values	90
5. Conclusions	92
References	94
Appendix A: Term by Term Integration of Formula (4.5) (Proof of) ..	97
Appendix B: Computer Programs (Descriptions and Listings)	100
B.1 General Programming Considerations	100
B.2 Subroutine SSYNTH	101
B.3 Subroutine HARMIN	106
B.4 Subroutine NORMAL	112
B.5 Subroutine NORMAX	124
B.6 Subroutine LEGFDN	131
B.7 Subroutine NVAR	134
B.8 Subroutine COVBLK	138

1. Introduction

Spherical harmonics are closely associated with the basic theory of gravitational and magnetic fields, such as those of the Earth and planets; for this reason they are important both in geodesy and in Earth and planetary physics.

The present work considers the numerical aspects of the reduction of global data sets to spherical harmonic coefficients, so the emphasis has been laid on the algorithms for this purpose. The procedures for harmonic analysis (and synthesis) given here are general enough to be used in the study of magnetic or electric fields, but most conclusions regarding their accuracy are restricted to the gravity field of our planet and to fields with the same power spectrum. The accuracy of these methods cannot be separated from the type of signal being used.

Modern instrumentation has provided scientists and engineers with vast amounts of information, and modern computers have made the processing of it possible, and even routine, thanks to constant improvements in both hardware and software. In the mid sixties, those branches of applied mathematics, physics, and engineering concerned with the sifting of data, or with the study of very large regular structures, were greatly affected by the advent of the Fast Fourier Transform (FFT). In spite of the fact that spherical harmonics are members of the family of Fourier transforms, closely related to two dimensional Fourier series, geodesy has lagged behind in the development of techniques similar to the FFT, partly because of the rather wicked nature of the sphere on which data are usually given, partly because there have not been enough data to make the development of powerful techniques a general concern. The topological differences between the Euclidean plane and the surface of the sphere may very well prevent the finding of algorithms as efficient as the FFT for the latter (certainly none seems to have been reported to date) but such algorithms should be regarded, nonetheless, as the desideratum for all those who wish to put their time and work into developing good numerical methods for spherical harmonic analysis.

The increasing use of artificial satellites for surveying the gravity field, particularly by radar altimetry and by the projected tracking of one satellite by another, are making the use of very efficient and accurate techniques for handling the resulting data indispensable: even a casual review of the literature of the last few years will show that serious efforts to provide such techniques are getting under way. The days of scarce, scattered, unreliable data are just about over.

The remainder of this section defines the basic problem and the associated notation, shows the relationships between spherical harmonics and 2D-Fourier series, presents some of the similitudes and differences between both, and explains some efficient algorithms for harmonic analysis and synthesis that are common to a number of different problems. Section 2 begins by defining a quadratic measure for the accuracy of the estimated harmonic coefficients based

on the covariance functions of the signal and the noise; a discussion on the optimization of this measure follows, leading to the application of least squares collocation to harmonic analysis. The use and implementation of least squares adjustment follows, and then a discussion of the connection between least squares and least squares collocation, shown as alternative and efficient techniques for solving the same problem. The section closes with an algorithm for the case when data are irregularly distributed. Section 3 illustrates with several numerical examples some of the methods presented earlier. Section 4 introduces an efficient formula for computing the covariances between block averages when analyzing mean values by collocation. This formula is much more efficient than others based on the numerical quadratures of the "point" covariance function.

1.1. Spherical Harmonic Analysis and Synthesis: Definitions

A square integrable, analytical function $f(\theta, \lambda)$ defined on the unit sphere $0 \leq \theta \leq \pi$ and $0 \leq \lambda \leq 2\pi$ can be expanded in a series of surface spherical harmonics

$$f(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{m=0}^n \bar{P}_{nm}(\cos \theta) [\bar{C}_{nm} \cos m\lambda + \bar{S}_{nm} \sin m\lambda] \quad (1.1)$$

where: \bar{P}_{nm} are the associated Legendre functions of the first kind, fully normalized so $\frac{1}{4\pi} \int_{\sigma} \bar{P}_{nm}(\cos \theta)^2 \left(\frac{\cos}{\sin}\right)^2 m\lambda d\sigma = 1$ (here $\int_{\sigma} d\sigma$ indicates integration on the unit sphere);
 $\bar{C}_{nm}, \bar{S}_{nm}$ are the fully normalized spherical harmonic coefficients.

For the sake of brevity, the following alternative notation shall be used when possible:

$$\bar{Y}_{nm}^{\alpha}(\theta, \lambda) = \begin{cases} \bar{P}_{nm}(\cos \theta) \cos m\lambda & \text{if } \alpha = 0 \\ \bar{P}_{nm}(\cos \theta) \sin m\lambda & \text{if } \alpha = 1 \end{cases}$$

and

$$\bar{C}_{nm}^{\alpha} = \begin{cases} \bar{C}_{nm} & \text{if } \alpha = 0 \\ \bar{S}_{nm} & \text{if } \alpha = 1 \end{cases}$$

The purpose of spherical harmonic analysis is to estimate the coefficients \bar{C}_{nm}^{α} from measurements of the signal $f(\theta, \lambda)$. These measurements, which may be corrupted by some noise or error signal "n", and which are assumed in what follows to be finite in number, constitute the data. The individual samples are called z_{ij} , so $z_{ij} = f(\theta_i, \lambda_j) + n_{ij}$. The subscripts i and j are used only to designate the position of the sample in a two-dimensional array, or grid, covering in some more or less regular way the sphere: i corresponds always to (co)latitude, and j to longitude. While, as in the last paragraph of section 3, some places in the grid may be empty, the grid itself is defined by a set of complete parallels and meridians. Unless otherwise specified, the separation between the lines of

latitude can be variable, but that between meridians is always constant and equal to $\Delta\lambda = \pi/N$, where N is an integer. The grid most often considered in this work is the equal angular grid, also called the regular grid; in this case $\Delta\theta$ is also constant and equal to $\Delta\lambda$.

For equal angular grids i and j take values $0 \leq i \leq N-1$ and $0 \leq j \leq 2N-1$, i increasing from North to South, and j from West to East. Formulas where the i, j subscripts appear are in the form appropriate to the regular grid, though most of them can be extended in a very simple way to other partitions.

Data may consist of values determined at the intersections of the grid, in which case they are referred to as "point data", or they may be averages over the blocks defined by the lines of the grid, and then they are called "area means" or "block means". In the equal angular grid the northernmost and southernmost blocks reach to the respective poles; there are N rows of blocks (i.e. blocks between the same parallels), there are $2N$ blocks per row, and i, j identify blocks according to the row and column they are in. Grids for equal angular point data may be "center point" grids, data being measured at the center of each block, so $i = 0$ corresponds to $\theta = \Delta\theta/2 = \Delta\lambda/2$.

No blocks stride the equator, so regular grids are symmetrical with respect to it; in other words: N is always even (extension to N odd is trivial).

Area means are identified here by overbars; they can be expanded in series simply by integrating (1.1) term by term, which can be done because the spherical harmonic series is always uniformly convergent for $0 \leq \theta \leq \pi$ and $0 \leq \lambda \leq 2\pi$:

$$\begin{aligned} \bar{f}_{ij} &= \frac{1}{\Delta_{ij}} \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{nm}^{\alpha} \int_{\sigma_{ij}} \bar{Y}_{nm}^{\alpha}(\theta, \lambda) d\sigma \\ &= \frac{1}{\Delta_{ij}} \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \int_{\theta_1}^{\theta_1 + \Delta\theta} \bar{P}_{nm}(\cos \theta) \sin \theta d\theta \left[\bar{C}_{nm}^{\alpha} \int_{\lambda_j}^{\lambda_j + \Delta\lambda} \cos m\lambda \right. \\ &\quad \left. + \bar{S}_{nm}^{\alpha} \int_{\lambda_j}^{\lambda_j + \Delta\lambda} \sin m\lambda d\lambda \right] \end{aligned} \quad (1.2)$$

Here \bar{f}_{ij} is the area mean of $f(\theta, \lambda)$ on the block σ_{ij} whose area is $\Delta_{ij} = \Delta\lambda(\cos \theta_1 - \cos(\theta_1 + \Delta\theta))$.

A basic property of spherical harmonics is their orthogonality:

$$\frac{1}{4\pi} \int_{\sigma} \bar{Y}_{nm}^{\alpha}(\theta, \lambda) \bar{Y}_{k,p}^{\beta}(\theta, \lambda) d\sigma = \begin{cases} 1 & \text{if } \alpha = \beta, m = p \text{ and } n = k \\ 0 & \text{otherwise} \end{cases} \quad (1.3)$$

as a consequence of which

$$\bar{C}_{nm}^{\alpha} = \frac{1}{4\pi} \int_{\sigma} \bar{Y}_{nm}^{\alpha}(\theta, \lambda) f(\theta, \lambda) d\sigma \quad (1.4)$$

Expression (1.4) is the inverse of (1.1); both constitute the basis of spherical harmonic analysis. In general, (1.4) cannot be calculated analytically, because what is available is not the function $f(\theta, \lambda)$, but a finite set of noisy measurements in the form of point values z_{ij} , or their area means \bar{z}_{ij} . Discretizing (1.4) on an equal angular grid results, for instance, in the following numerical quadratures formula:

$$\hat{C}_{nn}^\alpha = \frac{1}{4\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{Y}_{nn}^\alpha(\theta_i, \lambda_j) f(\theta_i, \lambda_j) \Delta_{ij} \quad (1.5)$$

where \hat{C}_{nn}^α indicates the estimate of \bar{C}_{nn}^α , as this type of formula is usually only an approximation. Formulas resembling (1.5) can be called "point values-type quadrature formulas", and can be handled by algorithms that are all identical from a structural point of view and whose prototype is that of paragraph 1.5.

If the data are block averages, several simple approximations have been proposed that take the form

$$\hat{C}_{nn}^\alpha = \mu_n \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{f}_{ij} \int_{\sigma_{ij}} \bar{Y}_{nn}^\alpha(\theta, \lambda) d\sigma$$

where the μ_n are scale factors. This kind of formula shall be studied further in section 2. Writing the above expression in full

$$\hat{C}_{nn}^\alpha = \mu_n \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{f}_{ij} \int_{\theta_i}^{\theta_i + \Delta\theta} \bar{P}_{nn}(\cos \theta) \sin \theta d\theta \int_{\lambda_j}^{\lambda_j + \Delta\lambda} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m\lambda d\lambda \quad (1.6)$$

This belongs to the general type

$$\hat{C}_{nn}^\alpha = K \sum_{i=0}^{N-1} \chi_i^{nn} \sum_{j=0}^{2N-1} \bar{f}_{ij} \left[\begin{Bmatrix} A(m) \\ -B(m) \end{Bmatrix} \cos mj\Delta\lambda + \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \sin mj\Delta\lambda \right] \quad (1.7)$$

where $B(m) = \begin{cases} (\cos m \Delta\lambda - 1)/m & \text{if } m \neq 0 \\ 0 & \text{if } m = 0 \end{cases}$; $A(m) = \begin{cases} (\sin m \Delta\lambda)/m & \text{if } m \neq 0 \\ \Delta\lambda & \text{if } m = 0 \end{cases}$

Formulas resembling (1.6) or (1.7) appear several times in this work, and are called here area means-type quadrature formulas.

If all \bar{C}_{nn}^α with $0 \leq n \leq N_{\max}$ are known, they can be used to compute the $(N_{\max} + 1)^2$ terms in

$$f(\theta, \lambda)_{N_{\max}} = \sum_{n=0}^{N_{\max}} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{nn}^\alpha \bar{Y}_{nn}^\alpha(\theta, \lambda) \quad (1.8)$$

which can be regarded as an approximation to $f(\theta, \lambda)$ at the point (θ, λ) . Expression (1.8), together with the truncation at N_{\max} of (1.2) (area means), defines the object of spherical harmonic synthesis: given the coefficients, estimate the function. As shown later, analysis and synthesis are related by a simple duality, so they require about the same number of operations when performed on a certain grid and with a given number of coefficients.

The set of all degree variances

$$\sigma_n^2 = \sum_{m=-n}^n \bar{C}_{nm}^2 + \bar{S}_{nm}^2 \quad (1.9)$$

constitutes the power spectrum of $f(\theta, \lambda)$. If all coefficients of degree $n > N_{\max}$ are zero, so are the corresponding σ_n ; in such case the function $f(\theta, \lambda)$ is said to be band limited.

Closely associated with the power spectrum is the isotropic covariance function

$$\text{cov}(f(P), f(Q)) = \sum_{n=0}^{\infty} \sigma_n^2 P_n(\cos \psi_{PQ}) \quad (1.10)$$

Here P and Q are two points on the sphere, $\psi_{PQ} = \cos^{-1}[\cos \theta_P \cos \theta_Q + \sin \theta_P \times \sin \theta_Q \cos(\lambda_P - \lambda_Q)]$ is the geocentric angle or spherical distance between both, and $P_n(\cos \psi)$ is the unnormalized Legendre polynomial of degree n . The power spectrum and the covariance of a function on the sphere, like those of functions defined on the real line, on the plane, and on Euclidean spaces of higher dimensions, can be used to obtain estimators related to discrete Wiener filters and predictors, i.e., satisfying a minimum variance condition. The spherical version of such technique is known as least squares collocation (Moritz, 1972). The inverse of expression (1.10) is

$$\sigma_n^2 = (2n+1) \int_0^\pi P_n(\cos \psi_{PQ}) \text{cov}(f(P), f(Q)) \sin \psi_{PQ} d\psi_{PQ} \quad (1.11)$$

Equations (1.10) and (1.11) show that, as usual, power spectrum and covariance function are linear transforms of each other. A formula such as (1.11) is sometimes called a Legendre transform.

In addition to the autocovariance (1.10), one can define, more generally,

$$\text{cov}(u(P), v(Q)) = \sum_{n=0}^{\infty} c_n^{u,v} P_n(\cos \psi_{PQ}) \quad (1.10^*)$$

as the covariance of two functions $u(\theta_P, \lambda_P)$ and $v(\theta_Q, \lambda_Q)$ on the sphere, where the

$$c_n^{u,v} = \sum_{m=-n}^n \bar{C}_{nm}^u \bar{C}_{nm}^v + \bar{S}_{nm}^u \bar{S}_{nm}^v, \quad n = 0, 1, \dots \quad (1.9^*)$$

constitute the "power crosspectrum".

A relationship similar to (1.11) applies to the $c_n^{u,v}$ and to $\text{cov}(u(P), v(Q))$.

A very important property of spherical harmonics is Parseval's theorem:

$$\frac{1}{4\pi} \int_{\sigma} f(\theta, \lambda)^2 d\sigma = \sum_{n=0}^{\infty} \sigma_n^2 \quad (1.12)$$

The orthogonality of spherical harmonics, and the fact that they form a complete orthogonal set of functions on the sphere, are among the reasons why they are used so widely in theory and in practice, but there is more to them than orthogonality.

The solid spherical harmonics

$$\frac{1}{r^{n+1}} \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda) \quad \text{and} \quad r^n \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda)$$

where r is the distance to the origin of coordinates, are all solutions of Laplace's equation, which in Cartesian coordinates is

$$\nabla^2 W = \frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2} + \frac{\partial^2 W}{\partial z^2} = 0 \quad (1.13)$$

This makes them appropriate for the study of harmonic functions, such as the gravitational potential, in spherical coordinates. Another property they have, unique among functions that are orthogonal on the sphere, is the relationship

$$\frac{1}{4\pi} \int_{\sigma} f(P) \sum_{n=0}^{\infty} (2n+1) k_n P_n(\psi_{PQ}) d\sigma = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 k_n \bar{C}_{n\alpha}^{\alpha} \bar{Y}_{n\alpha}^{\alpha}(Q) \quad (1.14)$$

which corresponds to the Convolution Theorem for ordinary Fourier series and is the basis of such fundamental formulas as Stokes' in gravimetric geodesy.

1.2. Relationship Between Spherical Harmonics and 2-D Fourier Series

As indicated in the previous paragraph, spherical harmonics share important properties with ordinary (trigonometric) Fourier series in one or more dimensions. There is also a very immediate relationship with ordinary two-dimensional (2-D) series that will be explained here. From Hobson (1931, Ch. III, formula (7)) we know that

$$P_{n\alpha}(\cos \theta) = \frac{(-1)^n (2n)!}{2^n n! (n-m)!} \sin^n \theta \left\{ \cos^{n-m} \theta - \frac{(n-m)(n-m-1)}{2(2n-1)} \cos^{n-m-2} \theta + \right. \\ \left. + \frac{(n-m) \dots (n-m-3)}{2 \cdot 4 \cdot (2n-1)(2n-3)} \cos^{n-m-4} \theta - \dots \right\} \quad (1.15)$$

so the normalized Legendre function is

$$\bar{P}_{n\alpha}(\cos \theta) = \frac{(-1)^n (2n)!}{2^n n! (n-m)!} \sqrt{\frac{2(2n+1)(n-m)!}{(m+n)!}} \sin^n \theta \sum_{k=0}^{L(n,m)} a_k(n,m) \cos^{n-m-2k} \theta$$

where

$$L(n,m) = \begin{cases} (n-m)/2 & \text{if } n-m \text{ is even} \\ (n-m-1)/2 & \text{if } n-m \text{ is odd} \end{cases}$$

while $a_k(n,m) = (-1)^k (n-m)(n-m-1) \dots (n-m-2k+1) [2 \cdot 4 \dots 2k(2n-1) \dots (2n-2k+1)]^{-1}$
for $k > 0$

In the interval $-\pi \leq \theta \leq \pi$ $\sin^m \theta$ is even when m is even, and odd when m is odd. In the interval $-\pi \leq \theta \leq \pi$ the sum $\sum_{k=0}^{l(n,m)} a_k(n,m) \cos^{n-m-2k} \theta$

is always even (it is a sum of even powers of $\cos \theta$), so the parity of $P_n(\cos \theta)$ is the same as that of m if $-\pi \leq \theta \leq \pi$. An even function can be expanded into a sum of cosines, and an odd function into a sum of sines. The highest frequency term will correspond to the highest frequency in the expansion of $\sin^m \theta \cos^{n-m} \theta$, so this term will be of the form $a_n \cos n\theta$ or $b_n \sin n\theta$. Therefore, the Legendre function satisfies one of the following equations in $-\pi \leq \theta \leq \pi$:

$$\bar{P}_{n,m}(\cos \theta) = \sum_{t=0}^n C_t^{n,n} \cos t\theta \quad (\text{a) if } m \text{ is } \underline{\text{even}};$$

$$\bar{P}_{n,m}(\cos \theta) = \sum_{t=0}^n S_t^{n,n} \sin t\theta \quad (\text{b) if } m \text{ is } \underline{\text{odd}}$$

A spherical harmonic $\bar{Y}_n^\alpha(\theta, \lambda)$ can have one of four possible forms:

$$\text{for } m \text{ } \underline{\text{even}}: \quad \bar{Y}_{n,n}^\alpha(\theta, \lambda) = \begin{cases} \sum_{t=0}^n C_t^{n,n} \cos t\theta \cos m\lambda & \text{for } \alpha = 0 \\ \sum_{t=0}^n C_t^{n,n} \cos t\theta \sin m\lambda & \text{for } \alpha = 1 \end{cases} \quad (1.16a)$$

$$\text{for } m \text{ } \underline{\text{odd}}: \quad \bar{Y}_{n,n}^\alpha(\theta, \lambda) = \begin{cases} \sum_{t=0}^n C_t^{n,n} \sin t\theta \cos m\lambda & \text{for } \alpha = 0 \\ \sum_{t=0}^n C_t^{n,n} \sin t\theta \sin m\lambda & \text{for } \alpha = 1 \end{cases} \quad (1.16b)$$

A sum of spherical harmonics such as (1.8) is equivalent to a sum of terms of the form $C_t^{n,n} \sin t\theta \cos m\lambda$, $C_t^{n,n} \sin t\theta \sin m\lambda$, $C_t^{n,n} \cos t\theta \cos m\lambda$ and $C_t^{n,n} \cos t\theta \sin m\lambda$, which are also the basic functions of 2-D Fourier series. The highest m and n in the spherical harmonic expansion

$$f(\theta, \lambda) = \sum_{n=0}^{N_{\max}} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{n,m}^\alpha \bar{Y}_{n,m}^\alpha(\theta, \lambda)$$

are equal to N_{\max} , so the highest degree and order (or spatial frequencies) in the Fourier series are also equal to N_{\max} . In conclusion: every surface spherical harmonic expansion where the highest degree is N_{\max} is identical to a 2-D Fourier series (where the highest m and n are also N_{\max}), in the domain $-\pi \leq \theta \leq \pi$, $0 \leq \lambda \leq 2\pi$. The converse is not true, because continuous functions on a sphere, such as the $\bar{Y}_{n,m}^\alpha$, must satisfy certain conditions at the poles that ordinary functions on the $-\pi \leq \theta \leq \pi$, $0 \leq \lambda \leq 2\pi$ domain do not have to. Spherical harmonics correspond to a subclass (linear subspace) of 2-D Fourier series.

For example: (Heiskanen and Moritz, Chapter I, 1967)

$$\bar{P}_{11}(\cos\theta) = \sqrt{3} \sin\theta$$

$$\bar{P}_{21}(\cos\theta) = \sqrt{30} \sin\theta \cos\theta = \frac{1}{2}\sqrt{30} \sin 2\theta$$

So

$$\bar{Y}_{11}^0 = \sqrt{3} \sin\theta \cos\lambda, \quad \bar{Y}_{11}^1 = \sqrt{3} \sin\theta \sin\lambda$$

$$\bar{Y}_{21}^0 = \frac{1}{2}\sqrt{30} \sin 2\theta \cos 2\lambda, \quad \bar{Y}_{21}^1 = \frac{1}{2}\sqrt{30} \sin 2\theta \sin 2\lambda$$

Calling the 2-D Fourier coefficients " a_{pn}^β ", where a_{pn}^0 correspond to terms of the form $\cos p\theta \cos m\lambda$

$$a_{pn}^1 \quad " \quad " \quad " \quad " \quad " \quad \cos p\theta \sin m\lambda$$

$$a_{pn}^2 \quad " \quad " \quad " \quad " \quad " \quad \sin p\theta \cos m\lambda$$

$$a_{pn}^3 \quad " \quad " \quad " \quad " \quad " \quad \sin p\theta \sin m\lambda$$

these can be related to the respective \bar{C}_{pn}^α by expressions of the form

$$\bar{C}_{pn}^\alpha = \frac{1}{4} \sqrt{\frac{2(2n+1)(n-m)!}{(n+m)!}} \sum_{p=0}^{n-1} I_{n,p}^\alpha a_{pn}^\beta \quad n = m, m+1, m+2, \dots \quad (1.17)$$

where $\beta = \alpha$ if m is even, and $\beta = \alpha + 2$ if m is odd. The $I_{n,p}^\alpha$ are defined as

$$I_{n,p}^\alpha = \int_0^\pi \cos p\theta P_n(\cos\theta) \sin\theta d\theta \quad \text{if } m \text{ is even}$$

$$I_{n,p}^\alpha = \int_0^\pi \sin p\theta P_n(\cos\theta) \sin\theta d\theta \quad \text{if } m \text{ is odd.}$$

and can be computed recursively using the formula

$$I_{n,p}^\alpha = \frac{2n-1}{2(n-m)} \left\{ I_{n-1,p+1}^\alpha + I_{n-1,p-1}^\alpha \right\} - \frac{n+m-1}{n-m} I_{n-2,p}^\alpha \quad (1.18a)$$

with the following starting values

$$I_{n,p}^\alpha = \begin{cases} 0 & \text{if } (m+p) \text{ is odd} \\ \frac{2(m+1)(2m)!}{2^n [(m+1)^2 - p^2] [(m-1)^2 - p^2] \dots [3^2 - p^2] [1^2 - p^2]} & \text{if } m \text{ is even, } p \text{ even,} \\ \frac{2(m+1)(2m)!}{2^n [(m+1)^2 - p^2] [(m-1)^2 - p^2] \dots [2^2 - p^2] [-p]} & \text{if } m \text{ is odd, } p \text{ odd.} \end{cases} \quad (1.18b)$$

The I_{np}^α are zero for alternate values of both p and m . These equations were reported by Ricardi and Burrows (1972), and show how to obtain the

$\bar{C}_{n,m}^{\alpha}$ once the $a_{n,m}^{\beta}$ have been computed from the data by means of the 2-D discrete Fourier transform. Normally this would be impossible, because the data only exists in the upper half of the $-\pi \leq \theta \leq \pi$, $0 \leq \lambda \leq 2\pi$ interval, and the 2-D algorithm requires information on all of it. However, the fact that the $\bar{P}_{n,m}(\cos \theta)$ are sums of sines only or cosines only makes the calculation possible.

While the maximum degree and order in a 2-D Fourier series do not reach the Nyquist frequency¹ ($m, n < N = \frac{\pi}{\Delta\lambda}$) in an equal angular grid, all the coefficients can be recovered exactly by solving $(2N)^2$ equations such as

$$f(\theta_1, \lambda_1) = \sum_{n=0}^N \sum_{m=0}^N \sum_{\beta=0}^3 a_{n,m}^{\beta} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}_n \theta_1 \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}_m \lambda_1 \quad (1.19)$$

When n or m exceed N_{max} , the matrix of the system of equations becomes singular, and the discrete Fourier transform consists of coefficients that "fit" the data, but differ from the true coefficients. The estimated coefficients are said to have been aliased with those that exceed the Nyquist frequency. In the case of spherical harmonics, which are a special case of 2-D Fourier series, a similar situation must arise: the harmonics in the data with $n \geq N$ are going to be aliased with those of lower degree, so the information available is not enough to recover all coefficients because the sampling is too coarse.

The aliasing of spherical harmonics sampled on regular grids is a consequence of the aliasing of the respective 2-D Fourier series, so it makes sense to talk of a "Nyquist frequency" in the case of those functions. Having established the connection between aliasing in both types of series, it is time to point out also some important differences.

1.3 Sampling Errors

Expressions (1.16a-b) shows that spherical harmonics are finite sums of 2-D Fourier harmonics, which is not the same as being each a Fourier harmonic. From this simple fact follow some important distinctions.

To understand them better, let us begin by stating some basic properties of Fourier series in one dimension, which carry over to higher dimensions but are easier to explain in one dimension.

If sampled at a constant interval $\Delta\lambda = \frac{\tau}{N}$, the following is always true of sines and of cosines:

¹Named after the Nyquist Theorem: the Fourier coefficients of a function of period $2N\Delta\lambda$ can be recovered only if $N_{max} < N$.

$$\left. \begin{aligned}
\sum_{k=0}^{2N-1} \cos mk\Delta\lambda \cos pk\Delta\lambda &= 0 & \text{if } m \neq p < N \\
\sum_{k=0}^{2N-1} \sin mk\Delta\lambda \sin pk\Delta\lambda &= 0 & \text{if } m \neq p < N \\
\sum_{k=0}^{2N-1} \cos mk\Delta\lambda \sin pk\Delta\lambda &= 0 & \text{for all } m \text{ and all } p.
\end{aligned} \right\} \quad (1.20-a)$$

These expressions are discrete counterparts of

$$\left. \begin{aligned}
\int_0^{2\pi} \cos m\lambda \cos p\lambda d\lambda &= \int_0^{2\pi} \sin m\lambda \sin p\lambda d\lambda = 0 \\
&\quad \text{when } m \neq p \\
\int_0^{2\pi} \cos m\lambda \sin p\lambda d\lambda &= 0 & \text{for all } m \text{ and } p.
\end{aligned} \right\} \quad (1.20-b)$$

and show that the orthogonal properties of sines and cosines are maintained when these are sampled regularly, provided the Nyquist frequency is not exceeded. From this follows that

$$a_m^Q = H \sum_{k=0}^{2N-1} \begin{pmatrix} \cos \\ \sin \end{pmatrix} mk\Delta\lambda f(k\Delta\lambda) = h \int_0^{2\pi} \begin{pmatrix} \cos \\ \sin \end{pmatrix} m\lambda f(\lambda) d\lambda \quad (1.21)$$

where $H = \begin{cases} \frac{1}{2N} & \text{if } m = 0 \\ \frac{1}{N} & \text{otherwise} \end{cases}$ and $h = \begin{cases} \frac{1}{2\pi} & \text{if } m = 0 \\ \frac{1}{\pi} & \text{otherwise} \end{cases}$, so the "Fourier"

counterpart of (1.5), (i.e. (1.21)) is an exact "numerical quadratures" formula for Fourier series. When the Nyquist frequency is exceeded, the trigonometric relationships

$$\cos mk\Delta\lambda = \cos(2N \pm m)k\Delta\lambda$$

$$\pm \sin mk\Delta\lambda = \sin(2N \pm m)k\Delta\lambda$$

imply that (1.21) will give, not the true coefficients, but the aliased estimates

$$\begin{aligned}
\hat{a}_m^0 &= a_m^0 + \sum_{h=0}^K a_{2hN+m}^0 + \sum_{h=1}^K a_{2hN-m}^0 \\
\hat{a}_m^1 &= a_m^1 + \sum_{h=0}^K a_{2hN+m}^1 - \sum_{h=1}^K a_{2hN-m}^1
\end{aligned} \quad (1.22)$$

with K such that $2KN$ is below, and $2(K+1)N$ is above the highest frequency present in $f(\lambda)$. Expression (1.20 a-b), (1.21), and (1.22) are the foundations of discrete Fourier analysis (also known as the computation of the discrete Fourier transform, or D.F.T.), and so well known that they are almost second nature to many engineers and scientists. Unfortunately, none of these discrete formulas has exact counterparts in spherical harmonic analysis, and this fact has been the cause of considerable confusion. The most common

misunderstandings seem to occur in the following areas:

(a) Number of equations and number of unknowns:

At each point in the grid it is possible to write an equation such as

$$f(\theta_i, \lambda_j) = \sum_{n=0}^{N_{\max}} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{nm}^{\alpha} \bar{Y}_{nm}^{\alpha}(\theta_i, \lambda_j) \quad (1.23)$$

similar to (1.19). The maximum number of coefficients that can be recovered without exceeding the Nyquist frequency (i.e., those with $n < N$) is N^2 . Coefficients of degree or order equal or higher than N shall not be, in general, free from aliasing. Since there are $2N^2$ points in an equiangular grid, it follows that the maximum number of fully recoverable coefficients is also half the number of points (equations) in the grid. By contrast, in 2-D Fourier analysis the number of coefficients equals the number of points in the grid (in the interval $-\pi \leq \theta \leq \pi$, $0 \leq \lambda \leq 2\pi$). One could have a grid with only N^2 points, as proposed by Giacaglia and Lundquist (1972), but such a grid would not be equal angular on the sphere.

The author had discussed this problem elsewhere (Colombo, 1979a, paragraph (5.2)), showing that the system of equations (1.23) becomes singular when all coefficients of degree and order $0 \leq (n, m) \leq M$ are included among the unknowns, and $M \geq N$. In other words: it is not possible to solve for a complete set of coefficients to degree and order $M \geq N$. The relevant part of that argument can be summarized as follows: the columns of A , the matrix of the system of equations (1.23), consist in successions of values $\bar{Y}_{nm}^{\alpha}(\theta_i, \lambda_j)$ of the harmonics corresponding to the unknowns \bar{C}_{nm}^{α} at the points (θ_i, λ_j) in the grid. The scalar product of two such columns is,

$$\sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{Y}_{nm}^{\alpha} \bar{Y}_{pq}^{\beta} = \sum_{i=0}^N \bar{P}_{nm}(\cos \theta_i) \bar{P}_{pq}(\cos \theta_i)$$

$$\times \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} q j \Delta \lambda = \begin{cases} 0 & \text{if } \alpha \neq \beta \\ 0 & \text{if } m \neq q \\ \neq 0 & \text{otherwise} \end{cases}$$

according to (1.20-a). Therefore, if two columns correspond to unknowns of different orders m and q , they must be orthogonal and, thus, independent. For the whole matrix not to be singular, all columns of the same order m must form sub-matrices $A(m)$ that have full rank. Otherwise, there will be columns in those $A(m)$, and consequently in A , that are linearly dependent, so A cannot be inverted. Consider $A(0)$, corresponding to all unknowns of order 0. This is a $2N^2 \times (M+1)$ matrix, and the elements of the columns of $A(0)$ have the same values as $\bar{P}_{n0}(\cos \theta_i)$, $0 \leq n \leq M$, $0 \leq i \leq N-1$. The \bar{P}_{n0} are functions of θ_i only, and there are N parallels in the grid, so there are no more than N independent rows in $A(0)$. Because the $\bar{P}_{n0}(\cos \theta)$ are polynomials of degree n in $\cos \theta$, these rows form a sub-matrix $S(0)$ of $A(0)$ that has $M+1$ independent columns,

as long as $M + 1 \leq N$ or $M < N$. If $M \geq N$ the extra columns will turn $S(0)$ into a rectangular matrix with more columns than rows; in other words: there will be no square submatrix in $A(0)$ of rank $M + 1$ if $M \geq N$, so $A(0)$ shall be rank deficient, and from this follows that A must be singular.

Emphasis must be placed on the word complete when referring to the set of "solvable" coefficients: it is possible, by removing some coefficients with $n < N$ from the unknowns, to introduce others in their stead with $n \geq N$, but then the solution will not be a complete set of coefficients.

(b) 100% aliasing:

If a term in a Fourier series has a frequency $n > N$, then it will be aliased with lower frequency terms and become impossible to discriminate from them. For most functions of practical interest, the higher the frequency, the smaller the term, so the coefficient estimated using the sum in (1.21) will be dominated by the lower frequency terms: the estimation error is thus likely to exceed 100%. Estimates above the Nyquist frequency are usually regarded as meaningless and the closer a term is to that frequency with increasing n , the less reliance is placed on its estimate.

In the case of spherical harmonics, expressions (1.16a-b) show clearly that the harmonic \bar{Y}_{nn}^α consists of several Fourier terms of frequencies ranging from 0 to n . When $n > N$, only that part of the Fourier expansion of \bar{Y}_{nn}^α above the Nyquist frequency will become scrambled beyond recovery; part of the harmonic is left intact: the low frequency "tail", which means that the effect of aliasing on the recovered coefficients does not necessarily reach 100% (or even 70%) at the Nyquist frequency, as shown in the examples of section 3.

(c) Orthogonality

From (1.20a) follows that the matrix of equations (1.19) for the Fourier series is orthogonal, so the coefficients estimated according to (1.21) are independent from each other.

In the case of the \bar{Y}_{nn}^α , orthogonality does not carry over to all the sampled harmonics, unless special "quadratures' weights" are introduced in (1.5) or (1.6). This lack of orthogonality affects, for instance, the formulas for mean values discussed in section 2. The method of Gaussian quadratures is an example of "quadratures with weights" that gives exact coefficients when the Nyquist frequency is not exceeded, though it requires a special grid where the parallels are situated at the same latitudes as the zeroes of $P_{n+1}(\cos \theta)$. The use of this method is possible because the product $[\bar{P}_{nn}(\cos \theta) \bar{P}_{pn}(\cos \theta)]$ is a polynomial in $\cos \theta$ of degree $n + p \leq 2N$; the grid, however, is an unusual one. Details of the application of Gaussian quadratures to spherical harmonic analysis are given in a report by Payne

(1971). Other examples of methods that recover the coefficients exactly when the data is noiseless and $N_{\alpha\alpha} < N$ are: least squares collocation, least squares adjustment, and the algorithm developed by Rice and Burrows from expressions (1.17) and (1.18 a-b).

In general, not all the discretised harmonics retain the orthogonality property, and the estimated coefficients are affected by the values of many of the other $C_{\alpha\alpha}$, in addition to those whose degrees exceed the Nyquist frequency. More about this will be said in section 2, when discussing least squares collocation and adjustment.

Summarising: there are enough differences between the aliasing of Fourier series and that of spherical harmonics, in spite of their being so closely related, to require a great deal of caution before using the intuition gained from one type of analysis when attempting the other. For this reason, the expression "aliasing error" will be replaced by the just as appropriate "sampling error", which is perhaps less charged with misleading connotations because it has not been applied almost exclusively to Fourier series.

It should be noticed here that Gaposchkin (1980) has published formulas for the sampling error on a type of equal area grid (i.e., all blocks have the same area as the equatorial ones). His formulas are the equivalent, for such a grid, of expression (1.22) for Fourier analysis, but much more complicated; they are made tractable numerically by the use of certain recursive expressions that he provides, thus showing an interesting new approach to the study of the problem.

1.4 Number of Operations in Analysis and in Synthesis

Expressions (1.17) and (1.18 a-b) can be used to calculate the $\hat{C}_{\alpha\alpha}$ once the 2-D Fourier coefficients of $f(\theta_1, \lambda_1)$, or $a_{p\alpha}^B$, have been estimated by means of the 2-D Discrete Fourier Transform (DFT). Using the Fast Fourier Transform (FFT) algorithm (Cooley and Tuckey, 1965), the number of operations required¹ is proportional to the number of data points, which is the order of N^2 , or $O(N^2)$ for short. The FFT is discussed further in paragraph 1.9.

Having obtained the $a_{p\alpha}^B$, calculation of (1.17) requires N operations per $\hat{C}_{\alpha\alpha}$, or $N \times N^2 = N^3$ for all of them. Finding the coefficients

¹ Most of these calculations have the form of scalar products

$$p = \sum_{k=0}^{N-1} a_k b_k, \text{ so the basic operation of finding } p^k = a_k b_k + p^{k-1} \text{ (} p^0 = 0, p^{N-1} = p \text{)}$$

 consists of one sum and one product.

$I_{n,p}^a$, by means of the recursives (1.18 a-b) adds another $O(N^3)$ operations that can be obviated by computing the $I_{n,p}^a$ once, and then storing them on magnetic tape or disk. One way or the other, $O(N^2) + O(N^3)$ operations are needed altogether, or $O(N^3)$ when N is large (say, $N \geq 180$).

As already mentioned, this procedure was first described by Ricardi and Burrows in 1972; more recently (1977) Goldstein developed a very similar idea and formulated a similar algorithm for synthesis. Goldstein's method uses recursive formulas for the $I_{n,p}^a$ different from (1.18 a-b). The synthesis algorithm also requires $O(N^3)$ operations.

As explained in paragraphs 1.5 through 1.7, the procedures presented there also require $O(N^3)$ operations for analysis, and as many for synthesis, though they are formally different from Ricardi and Burrows'. In paragraph 1.8 it will be shown that synthesis requires as many operations as analysis, because one is the dual of the other.

The fact that two rather different approaches (Ricardi and Burrows' and the one described in this report) require essentially the same number of operations suggests that " $O(N^3)$ " might be a property of all analysis and synthesis algorithms on regular spherical grids due, somehow, to the nature of the sphere itself. This is speculation, of course, but if not, are there other ways of partitioning the sphere for which faster methods exist? The author has discussed this possibility before (Colombo, 1979, paragraph 4.6). It is interesting to notice that in all these procedures the $O(N^3)$ operations are those associated with θ_1 , or "column operations"; "row operations" are only $O(N^2)$. In the case of the Euclidean plane, the 2-D Fourier transform requires the same number of operations per row than per column, $O(N)$, thus the total is only $O(N^2)$, or $O(N)$ times faster than its spherical "counterpart."

While not as efficient as the 2-D FFT, the algorithms for the sphere considered here can be much faster than the straightforward implementation of expressions (1.5), (1.6), (1.2), or (1.8). The latter has been the approach of many scientists who have developed their own software, but whose main interest has generally been far removed from the study of numerical techniques. In 1976, while working at the University of New South Wales (Australia), the author developed the two algorithms of paragraphs 1.5 and 1.6, and C. Rizos programmed them. Subsequently they were used at Goddard Space Flight Center, in Maryland. To everybody's surprise, Rizo's programs turned out to be more than 100 times faster than those in use at the time, when run under the same conditions. More recently, this author has written the subroutines HARMIN and SSYNTH described in appendix B. SSYNTH has been used, after the fashion of the numerical experiments described in Section 2, to generate 64000 $1^\circ \times 1^\circ$ mean values (simulated averaged gravity anomalies), each the sum of the 90000 terms of an expansion

complete to degree and order 300. This took less than 50 central processor unit seconds in the AMDHAL 470 V/6-II owned by The Ohio State University (OSU). All calculations were in double precision (32 bits words), using FORTRAN H EXTENDED. Precomputed values of the Legendre functions (or their integrals) were read from tape, and all operations involving trigonometric functions were carried out by a Fast Fourier Transform subroutine. These two characteristics, plus a generally tighter coding, are the reasons for the greater speed of this program, compared to the older versions mentioned before.

In all these methods all operations along a given row or parallel (constant θ_1) are independent from those for any other row, so a parallel processing computer with N processors (arithmetic and control units) could analyse or synthesize a full grid of N rows as fast as an ordinary computer with one central processor can do a single row. This N-fold increase in speed can be obtained with the same type of basic hardware (gates, registers) that is currently used in conventional "general purpose" main-frame machines. The full power of the algorithms presented in this work will be realized when computers of parallel structure become more widely available for scientific applications than they are today.

1.5 Algorithm for the Analysis of Point Values

Expression (1.5) written in full becomes

$$\hat{C}_{nn}^{\alpha} = \frac{1}{4\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{P}_{nn}(\cos \theta_1) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_1, \lambda_j) \Delta_{1j}$$

which corresponds to the general type

$$\hat{C}_{nn}^{\alpha} = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} x_i^{n\alpha} \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_1, \lambda_j) \Delta_{1j} \quad (1.24)$$

where $x_i^{n\alpha}$ could be $\frac{1}{4\pi} \bar{P}_{nn}(\cos \theta_1) \Delta_{1j}$ as above, or $\bar{P}_{nn}(\cos \theta_1) \omega_i$ in the case of quadrature with weights ω_i , etc.

To simplify the discussion, the grid is supposed to be equal angular and $N_{\alpha\alpha} = N-1$. This and the algorithms that follow can be easily adapted for the equal area grids current today. Subroutines HARMIN AND SSYNTH (Appendix B) can handle the cases $N_{\alpha\alpha} < N-1$ and $N_{\alpha\alpha} > N-1$ as well as $N_{\alpha\alpha} = N-1$.

The equal angular grid is symmetrical with respect to the Equator, and assuming that (the same as $\bar{P}_{nn}(\cos \theta_1)$ or $\bar{P}_{nn}(\cos \theta_1) \sin \theta_1$) $x_i^{n\alpha} = x_{N-1-i}^{n\alpha}$ if $n-m$ is even, and $x_i^{n\alpha} = -x_{N-1-i}^{n\alpha}$ if $n-m$ is odd, one can write

$$\begin{aligned} \hat{C}_{n,m}^{\alpha} = & \sum_{i=0}^{\frac{1}{2}N-1} \left(\chi_i^{n,m} \left[\sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_i, \lambda_j) \right] + \right. \\ & \left. (-1)^{n-m} \chi_i^{n,m} \left[\sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_{N-1-i}, \lambda_j) \right] \right) \end{aligned} \quad (1.25)$$

This formula suggests the following procedure in two nested loops:

START: set $i = -1$, $\hat{C}_{n,m}^{\alpha(0)} = 0$ for all $0 \leq n, m \leq N$;

Outer loop:

(a) increment i by 1 unless $i = \frac{1}{2}N - 1$, in which case STOP

Inner loop:

(b) compute all

$$\begin{aligned} \begin{Bmatrix} a_i^1 \\ b_i^1 \end{Bmatrix} &= \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_i, \lambda_j) \\ \begin{Bmatrix} a_{N-1-i}^{N-1} \\ b_{N-1-i}^{N-1} \end{Bmatrix} &= \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda f(\theta_{N-1-i}, \lambda_j) \end{aligned}$$

(c) find all

$$\hat{C}_{n,m}^{\alpha(i)} = \hat{C}_{n,m}^{\alpha(i-1)} + K \left[\begin{Bmatrix} a_i^1 \\ b_i^1 \end{Bmatrix} + (-1)^{n-m} \begin{Bmatrix} a_{N-1-i}^{N-1} \\ b_{N-1-i}^{N-1} \end{Bmatrix} \right] \chi_i^{n,m} \quad (1.26)$$

for $0 \leq n, m \leq N$ (where " $(-1)^{n-m} \{ \}$ " merely indicates that " $\{ \}$ " is to be added or subtracted according to the parity of $(n-m)$); GO BACK TO (a).

At the end of the outer loop, $\hat{C}_{n,m}^{\alpha(\frac{1}{2}N-1)} = \hat{C}_{n,m}^{\alpha}$.

The a_i^1 and b_i^1 in (1.26) can be computed by taking the Fourier transform along row i and row $N-1-i$ of the values of $f(\theta, \lambda)$. This involves $O(N)$ operations. There are half as many $\chi_i^{n,m}$ as $\hat{C}_{n,m}^{\alpha(i)}$, therefore it takes $2(N+1)^2$ products, and just as many sums, to form the $\hat{C}_{n,m}^{\alpha(i)}$ for the pair of rows i and $N-1-i$. Consequently, there are $O(N) + O(N^2)$ operations per pair of rows, or $O(N^2) + O(N^3)$ for the grid as a whole. This is the same as with the Ricard and Burrows' algorithm, quickly approaching $O(N^3)$ as N increases.

Subroutine HARMIN (Appendix B) implements this technique.

1.6 Algorithms for the Synthesis of Point Values

Expression (1.8) is of the general form

$$f(\theta_i, \lambda_j) = \sum_{n=0}^{N-1} \sum_{m=0}^n \chi_i^{nm} [\bar{C}_{nm} \cos m \lambda_j + \bar{S}_{nm} \sin m \lambda_j] \quad (1.27)$$

where χ_i^{nm} can be, for instance, $K \bar{P}_{nm}(\cos \theta_i)$, etc., with K being a proportionality constant. Rearranging terms and considering the parity of χ_i^{nm} leads to

$$\begin{aligned} \left\{ \begin{array}{l} f(\theta_i, \lambda_j) \\ f(\theta_{N-1-i}, \lambda_j) \end{array} \right\} &= \sum_{n=0}^{N-1} \left(\left[\sum_{m=n}^{N-1} \left\{ \chi_i^{nm} \right. \right. \right. \\ &\quad \left. \left. \left. (-1)^{n-m} \chi_i^{nm} \right\} \bar{C}_{nm} \right] \cos m j \Delta \lambda + \right. \\ &\quad \left. \left[\sum_{m=n}^{N-1} \left\{ \chi_i^{nm} \right. \right. \right. \\ &\quad \left. \left. \left. (-1)^{n-m} \chi_i^{nm} \right\} \bar{S}_{nm} \right] \sin m j \Delta \lambda \right) \end{aligned} \quad (1.28)$$

which suggests a procedure in two nested loops:

START: set $i = -1$

Outer loop:

(a) increment i by 1, unless $i = \frac{1}{2}N - 1$, in which case STOP

Inner loop:

(b) compute all

$$\begin{aligned} \left\{ \begin{array}{l} \alpha_n^i \\ \beta_n^i \end{array} \right\} &= \sum_{m=n}^{N-1} \chi_i^{nm} \left\{ \begin{array}{l} \bar{C}_{nm} \\ \bar{S}_{nm} \end{array} \right\} \\ \left\{ \begin{array}{l} \alpha_n^{N-1-i} \\ \beta_n^{N-1-i} \end{array} \right\} &= \sum_{m=n}^{N-1} \chi_i^{nm} (-1)^{n-m} \left\{ \begin{array}{l} \bar{C}_{nm} \\ \bar{S}_{nm} \end{array} \right\} \end{aligned}$$

for $0 \leq m \leq N$;

(c) find all

$$\begin{aligned} \left\{ \begin{array}{l} f(\theta_i, \lambda_j) \\ f(\theta_{N-1-i}, \lambda_j) \end{array} \right\} &= K \sum_{n=0}^{N-1} \left(\left\{ \begin{array}{l} \alpha_n^i \\ \alpha_n^{N-1-i} \end{array} \right\} \cos m j \Delta \lambda + \right. \\ &\quad \left. \left\{ \begin{array}{l} \beta_n^i \\ \beta_n^{N-1-i} \end{array} \right\} \sin m j \Delta \lambda \right) \end{aligned} \quad (1.29)$$

for $0 \leq j \leq 2N - 1$ (where $(-1)^{n-m} \{ \dots \}$ means the same as in the previous paragraph); GO BACK TO (a).

At the end of the outer loop all $f(\theta_i, \lambda_j)$ in the grid are known.

Expression (1.29) is computed by applying the FFT to the $2N$ α_n^1, β_n^1 (per row) and the $2N$ $\alpha_n^{N-1}, \beta_n^{N-1}$, taking $O(N)$ operations. The first part of the inner loop (forming the α_n^1, β_n^1 and $\alpha_n^{N-1}, \beta_n^{N-1}$) involves $O(N^2)$ operations; for all $\frac{1}{2}N$ pairs of rows the total is $O(N^2) + O(N^3)$, and this tends to $O(N^3)$ as N increases.

Subroutine SSYNTH (Appendix B) implements this technique.

1.7 Algorithms for the Analysis and Synthesis of Area Means

(I) Analysis

Rearranging (1.7)

$$\hat{C}_{n,n}^{\alpha} = \sum_{i=0}^{N-1} \chi_i^{n,n} \left[\frac{A(m)}{B(m)} (a_n^1 + (-1)^{n-n} a_{n-1}^{N-1}) + \frac{B(m)}{A(m)} (b_n^1 + (-1)^{n-n} b_{n-1}^{N-1}) \right] \quad (1.30)$$

where a_n^1, b_n^1 are the same as in paragraph (1.5). A procedure similar to that for the analysis of point values can be obtained directly by replacing the bracket in (1.26) with that in (1.30), and then proceeding as in the algorithm for point values. The total number of operations is, once more, $O(N^2) + O(N^3)$, or $O(N^3)$ for large N . Subroutine HARMIN also implements this algorithm.

(II) Synthesis

Truncating the series in equation (1.2), replacing $\int_{\theta_1}^{\theta_1 + \Delta\theta} \bar{P}_{n,n}(\cos\theta) \sin\theta d\theta$ with $\chi_i^{n,n}$, rearranging terms, considering the parity of $\chi_i^{n,n}$, and using α_n^1, β_n^1 as defined in paragraph (1.6), leads to the expression

$$\left\{ \frac{\bar{F}_{1,j}}{\bar{F}_{N-1-j,j}} \right\} = \sum_{n=0}^{N-1} \left[\left\{ \frac{\alpha_n^1}{\alpha_n^{N-1}} \right\} A(m) - \left\{ \frac{\beta_n^1}{\beta_n^{N-1}} \right\} B(m) \right] \cos m j \Delta\lambda + \left[\left\{ \frac{\alpha_n^1}{\alpha_n^{N-1}} \right\} B(m) + \left\{ \frac{\beta_n^1}{\beta_n^{N-1}} \right\} A(m) \right] \sin m j \Delta\lambda \quad (1.31)$$

The algorithm for the synthesis of the $\bar{F}_{1,j}$ is a direct extension of that for point values. The number of operations, once more, is $O(N^2) + O(N^3)$ for large N . Subroutine SSYNTH implements this algorithm as well.

1.8 Duality between Analysis and Synthesis

Pairs of direct and inverse linear transforms, such as Fourier transforms, possess dual characteristics: certain words and mathematical

expressions can be arranged in pairs (a,b) such that, if every "a" is replaced by its "b" in any statement or equation valid for a function f, then the modified statement is valid for the transform F of f, and viceversa.

Analysis and synthesis of spherical harmonics are reciprocal linear transformations of data into coefficients and of coefficients into "data" closely akin to Fourier transforms, so they can be expected to exhibit some dual properties. Comparing the formulas and algorithms in paragraphs (1.5), (1.6) and (1.7) shows many similarities, among them the number of operations. This can be understood as being a consequence of duality. To make this point clear, consider the following pairs:

$$\left(\begin{Bmatrix} f(\theta_1, \lambda_j) \\ f(\theta_{N-1}, \lambda_j) \end{Bmatrix}, \begin{Bmatrix} \bar{C}_{nn} \\ \bar{S}_{nn} \end{Bmatrix} \right); \left(\begin{Bmatrix} \hat{C}_{nn} \\ \hat{S}_{nn} \end{Bmatrix}, \begin{Bmatrix} f(\theta_1, \lambda_j) \\ f(\theta_{N-1}, \lambda_j) \end{Bmatrix} \right) \text{ (and similarly with } \begin{Bmatrix} \bar{I}_{1j} \\ \bar{I}_{N-1,j} \end{Bmatrix} \text{)};$$

$$(\chi_1^{nn} (-1)^{n-n}, \sin mj\Delta\lambda); (i, m); (j, n); (\chi_1^{nn}, \cos mj\Delta\lambda)$$

$$\left(\sum_{i=0}^{N-1}, \sum_{n=0}^{N-1} \right); \left(\sum_{j=0}^{N-1}, \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \right)$$

From these we can derive the following pairs:

$$\left(\begin{Bmatrix} a_n^i \\ a_{N-1-i}^i \end{Bmatrix}, \begin{Bmatrix} \alpha_n^i \\ \alpha_{N-1-i}^i \end{Bmatrix} \right); \left(\begin{Bmatrix} b_n^i \\ b_{N-1-i}^i \end{Bmatrix}, \begin{Bmatrix} \beta_n^i \\ \beta_{N-1-i}^i \end{Bmatrix} \right); \left(\begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix}, \begin{Bmatrix} \alpha_n^i \\ \alpha_{N-1-i}^i \end{Bmatrix} \right); \left(\begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix}, \begin{Bmatrix} \beta_n^i \\ \beta_{N-1-i}^i \end{Bmatrix} \right);$$

$$(a_n^i, A(m)); (b_n^i, B(m)); (a_{N-1-i}^i, -B(m)); (b_{N-1-i}^i, A(m)); \text{ etc.,}$$

and, in conclusion:

("ANALYSIS", "SYNTHESIS")

Each one of the analysis algorithms becomes its synthesis counterpart by a simple replacement of terms. Once an algorithm for analysis (synthesis) is defined, the corresponding algorithm for synthesis (analysis) follows. For instance, one can easily apply the principle of duality to the Ricardí and Burrows' method of paragraph (1.4) to obtain a synthesis technique.

1.9 Usefulness of the Fast Fourier Transform Method

The excellent book by Brigham (1974) gives a thorough presentation of the 1-D discrete Fourier transform and its applications, and explains in detail the method known as FFT for computing such transform. The Fourier transform in 2 and higher dimensions can be found simply as follows:

first, get the 1-D transform of each row, then that of each column of the modified array . . . , etc., until all dimensions have been exhausted in this way. Understanding the workings of the 1-D transform is enough to understand those of the N-dimensional transform as well. The FFT requires $O(\text{number of points})$ operations for each row along the n th dimension, so the number for all points in a regular, euclidean array is always of the order of the total number of points in that array.

Before the mid-sixties _ when the FFT came along _ the best techniques available for the analysis of data on regular arrays required $O(\text{number of points})^2$ operations. The increase in speed of $O(\text{number of points})$ brought about a true revolution in data processing: work that had been long regarded as impossible became feasible overnight, the field of industrial and scientific applications for numerical Fourier transforms expanded tremendously; the impact in areas as diverse as cristallography and communications engineering was remarkable.

Having mentioned the positive side of the FFT, which is used in the algorithms described so far (at least in principle) and in the programs HARMIN and SSYNTH, it is only proper to say something about its alternatives. The FFT calculates all $2N$ Fourier coefficients a_n, b_n very efficiently, but takes just about as many operations to get only a few coefficients as it takes to get all: for N_{max} small compared to N there may be a real disadvantage in using the FFT. The FFT is most efficient when the grid is such that N is an integer power of 2. The grids used in geodesy are usually based on the division of the circle in 360° , and many on the sexagesimal division of 1° as well. In all of these N contains factors other than 2, so a less efficient version of the FFT, known as the mix-radix FFT (Singleton's algorithm) must be used.

Finally, the mix-radix algorithm is rather convoluted, so it is best to take ready available subroutines from software libraries (as it is done in HARMIN and SSYNTH) rather than to incorporate the FFT "on line" in the program one is writing. This means that the program is going to be less self-contained.

The "pre-FFT" methods can be more efficient than the FFT when $N_{\text{max}} \ll N$; they are also very easy to program. For the sake of completeness, the outline of a method this author has used quite often will be given here.

Consider the trigonometric relationships

$$\cos(\alpha \beta) = 2\cos \beta \cos(\alpha - 1) \beta - \cos(\alpha - 2) \beta \quad (1.32)$$

$$\sin(\alpha \beta) = 2\cos \beta \sin(\alpha - 1) \beta - \sin(\alpha - 2) \beta \quad (1.33)$$

with α, β real. If all the values of the trigonometric functions could be obtained with one operation each¹, the number of operations involved in finding $\begin{Bmatrix} a_m \\ b_m \end{Bmatrix} = \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m j \Delta \lambda (\lambda_j) f(\theta_1, \lambda_j)$ would be $2(2N)$, or $4N^2$ for

all $0 \leq m \leq N$. In other words: $O(N^2)$, as expected. This is precisely what can be done using (1.32), (1.33) as recursive expressions for $\cos m(j\Delta\lambda)$ and $\sin m(j\Delta\lambda)$ with m integer and $0 \leq j \leq 2N-1$. The values of $\cos m(-\Delta\lambda) = \cos m\Delta\lambda$ and $\sin m(-\Delta\lambda) = -\sin m\Delta\lambda$, are needed to start the recursion; they can be calculated with standard trigonometric subroutines. The use of such subroutines increases the number of operations slightly over $4N^2$, but if N is large enough, this is negligible.

With all calculations carried out in double precision, this method gives values of cosine and sine that coincide to better than 5 significant figures with those provided by the standard FORTRAN functions, when N is as large as 1800 ($0.1^\circ \times 0.1^\circ$ grid). By taking advantage of half-wave symmetries in the sine and cosine, and by ingenious programming, the number of operations can be reduced by a factor of 4 or more rather easily.

1.10 Functions Harmonic in Space and their Gradients

If $f(\theta, \lambda, r)$ satisfies Laplace's equation $\nabla^2 f = 0$ in the space outside a sphere of radius a , then it can be represented, in that space, by the solid spherical harmonic expansion

$$f(\theta, \lambda, r) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \frac{a^n}{r^{n+1}} \bar{C}_{n\alpha}^{\alpha} \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda) \quad (1.34)$$

If we consider, at a point $P \equiv (\theta, \lambda, r)$ in space, the local triad $\vec{r}, \vec{h}, \vec{t}$ oriented downwards to the origin, West to East along the tangent to the local parallel, and North-South along the local meridian, the components of the gradient of $f(\theta, \lambda, r)$ along this three axes are

$$\frac{\partial f}{\partial r}(\theta, \lambda, r) = \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \frac{a^n}{r^{n+2}} (n+1) \bar{C}_{n\alpha}^{\alpha} \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda) \quad (1.35)$$

$$\frac{\partial f}{\partial h}(\theta, \lambda, r) = \sum_{n=0}^{\infty} \sum_{m=0}^n \frac{a^n}{r^{n+2}} \bar{P}_{n\alpha}(\cos \theta) \operatorname{cosec}(\theta) [m \bar{S}_{n\alpha} \cos m\lambda - m \bar{C}_{n\alpha} \sin m\lambda] \quad (1.36)$$

$$\frac{\partial f}{\partial t}(\theta, \lambda, r) = \sum_{n=0}^{\infty} \sum_{m=0}^n \frac{a^n}{r^{n+2}} \frac{d\bar{P}_{n\alpha}}{d\theta}(\cos \theta) [\bar{C}_{n\alpha} \cos m\lambda + \bar{S}_{n\alpha} \sin m\lambda] \quad (1.37)$$

¹Here one "operation" as mentioned in paragraph (1.4), consists of one sum and one product.

Expressions (1.34), (1.35), (1.36), and (1.37) appear often in the discussion of geodetic problems, and their calculation from a set of coefficients is an important problem. If all the values of any of them (with r constant, and on a regular grid) are required, the methods discussed so far can be used after a few minor additions.

The synthesis algorithms can be thought of, for the purpose of this discussion, as "black boxes" with the coefficients $\bar{C}_{n,m}^\alpha$, the Legendre functions, $N_{n,m}$, and N as inputs, and all the $2N^2$ values of $f(\theta, \lambda)$ on the corresponding regular grid as outputs. To compute the expressions given above, only the part of the input consisting in the coefficients and/or the Legendre functions has to be modified before they enter the "box", which remains untouched. For instance: to compute (1.34) one should replace " $\bar{C}_{n,m}^\alpha$ " with " $\frac{a_n}{r^{n+2}} \bar{C}_{n,m}^\alpha$ " in the "input"; the others are equally obvious and will not be explained further. The following recursive formulas can be used to obtain the derivatives of the Legendre functions:

$$\frac{d\bar{P}_{n,m}}{d\theta} = (\sin\theta)^{-1} \left\{ n \bar{P}_{n,m}(\cos\theta) \cos\theta - \left[\frac{(n^2 - m^2)(2n+1)}{(2n-1)} \right]^{-\frac{1}{2}} P_{n-1,m}(\cos\theta) \right\} \quad (1.38a)$$

$$\frac{d\bar{P}_{n,0}}{d\theta} = \left[\frac{(2n+1)}{(2n)} \right]^{\frac{1}{2}} \left\{ \sin\theta \frac{d\bar{P}_{n-1,n-1}}{d\theta} + \cos\theta P_{n-1,n-1} \right\} \quad (1.38b)$$

with the starting value

$$\frac{d\bar{P}_{0,0}}{d\theta} = 0$$

These recursives follow from the unnormalized formula

$$(\cos^2\theta - 1) \frac{dP_{n,m}(\cos\theta)}{d(\cos\theta)} = n \cos\theta P_{n,m}(\cos\theta) - (n+m) P_{n-1,m}(\cos\theta)$$

(N.N. Lebedev, "Special Functions", Dover, 1972, Ch. 7, equation 7.12.16), and from

$$\bar{P}_{n,m} = \left[\frac{(2n+1)}{(2n-1)} \right]^{\frac{1}{2}} \sin\theta \bar{P}_{n-1,n-1} \quad (\text{see paragraph (4.4)})$$

and

$$\bar{P}_{n,n} = \left[\frac{(n+m)!}{2(2n+1)(n-m)!} \right]^{-\frac{1}{2}} P_{n,n}; \quad \bar{P}_{n,0} = \sqrt{2n+1} P_{n,0}$$

The complete recursive expressions for the $\bar{P}_{n,m}$ are given in paragraph (4.4).

The expansion for the area means defined by (1.2) can be differentiated term by term because it converges uniformly. The expressions for area means gradients, equivalent to those given here for point values, are immediate. They can be computed after simple modifications to the $\bar{C}_{n,m}^\alpha$ and/or the $\chi_1^{n,m}$, and using the same programs for computing the area means.

Subroutine "LEGFDN", listed on Appendix B, can compute both the normalized $\bar{P}_{n,m}(\cos\theta)$ and their derivatives $\frac{d\bar{P}_{n,m}}{d\theta}(\cos\theta)$.

2. Error Measure and Optimal Quadrature Formulas

This section introduces a criterion for quantifying the errors of numerical quadratures formulas that is based on the statistical properties of the data. Three qualities are highly desirable in an error measure: (a) it should be easy to determine; (b) it should be mathematically tractable; (c) it should provide a good idea of the likely size of the actual errors. Point (a) is taken into account by choosing a quadratic measure, because the numerical formulas are linear estimators of the \bar{C}_{nn}^α , and the linear, quadratic estimation problem is fairly simple, with its mathematical side very well understood and developed today, which takes care also of (b). Regarding (c), the reader will have to wait till section three, where certain evidence, obtained from numerical experiments, supports the assertion that, though statistical in nature, the error measure adopted represents the actual errors very closely.

Having defined the error measure, the notion of optimal or best formula according to such measure is investigated, leading to the application of least squares collocation and least squares adjustment to spherical harmonic analysis.

2.1 The Isotropic Covariance

The isotropic covariance (expression (1.10)) between two functions $u(\theta, \lambda)$, $v(\theta, \lambda)$ on the unit sphere, both expandible in spherical harmonic series

$$u(\theta, \lambda) = K_u \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{nn}^u \bar{Y}_{nm}^\alpha(\theta, \lambda);$$

$$v(\theta, \lambda) = K_v \sum_{n=0}^{\infty} \sum_{m=0}^n \sum_{\alpha=0}^1 \bar{C}_{nn}^v \bar{Y}_{nm}^\alpha(\theta, \lambda)$$

can be formally defined as follows

$$\text{cov}(u(P), v(Q)) = M \{ u(P) v(Q) \} \quad (2.1)$$

where $M \{ \}$ is the isotropic averaging operator and P and Q are two points on the sphere separated by the spherical distance ψ_{PQ} . The operator $M \{ \}$ symbolizes the average of its argument (in the present case the product $u(P) v(Q)$) over all rotations of the sphere. This can be visualized if one thinks of the points P and Q as given in a fixed system of coordinates, while the sphere, on which u and v are defined, rotates in all possible ways. After all the (infinitely many) possible rotations, the average product $u(P) v(Q)$ will be identical to $\text{cov}(u(P), v(Q))$. This kind of covariance, though purely geometrical, resembles closely that of stochastic processes such as time series.

The importance of the isotropic operator and the isotropic covariance function in spherical harmonic analysis stems from the fact that the latter can be described as the estimation of certain parameters of a function $f(\theta, \lambda)$, the \bar{C}_{nn}^α , from data sampled on a sphere. From paragraph (2.7) on, this report deals with optimal estimators for the \bar{C}_{nn}^α based on the theory of least squares collocation. Such optimal estimators minimize a quadratic measure of the error that is defined in terms of the operator $M \{ \}$, this measure being introduced in paragraph (2.4).

The idea of least squares collocation is related to the basic principles of such linear, minimum variance estimators for time series as the Wiener and Kalman filters, which have found wide application in the physical sciences and in engineering over the last thirty years, and have been generalized to deal with both continuous and discrete time processes, and also "processes" in more than one dimension, such as are found in pattern recognition and in digital image enhancement. Two-dimensional Wiener filtering, of which the reader can find several fine descriptions in the special issue of the Proceedings of the IEEE, Vol 65, No. 6, 1977, is also applicable to "flat-Earth" geodetic calculations; least squares collocation can be regarded as the extension of this type of filtering to calculations on the sphere. Isotropic average operators are not the only ones that could be used in the "statistical" approach, though they are probably the easiest to work with and, perhaps, the best for the sort of application considered here. For a description of other likely operators, the reader is referred to the paper by Rummel and Schwarz (1978). Probably the most didactic introduction to the method of collocation remains Heiskanen and Moritz, (Ch. 7, 1967).

Reasoning as in Heiskanen and Moritz (ibid), one can show that

$$\text{cov}(u(P), v(Q)) = \sum_{n=0}^{\infty} c_n^{u,v} P_n(\cos \psi_{PQ})$$

which is, in fact, expression (1.10*) _the definition of the isotropic covariance given in section 1 without any reference to $M \{ \}$. Similarly,

$$\text{cov}(u(P), u(Q)) = \sum_{n=0}^{\infty} \sigma_n^2 P_n(\cos \psi_{PQ})$$

usually known as "the covariance of u" (expression 1.10), while (1.10*) represents the "covariance between u and v", or "the crosscovariance of u and v". The one-to-one relationship between covariance and power spectrum (or crosscovariance and crosspectrum) should be clear from these expressions.

To apply the notions introduced above to the \bar{C}_{nn}^α , it is necessary to think of them as functions rather than fixed values. This is possible if one considers changes in the coordinates θ, λ brought about by rotation. Each

such change results in different coefficients, though the function they describe is the same, only rotated with respect to the old system. The new system can be related to the first by three angles: the coordinates θ, λ of the shifted north pole, and the azimuth A of the zero meridian. Therefore, the $\bar{C}_{n, m}^\alpha$ are functions of θ, λ, A and this is enough to define the average over all rotations of the product of each $\bar{C}_{n, m}^\alpha$ by itself or by another function, in a meaningful way. Two important properties of spherical harmonics are:

$$(A) \quad M \left\{ \sum_{m=0}^n \bar{C}_{n, m}^2 + \bar{S}_{n, m}^2 \right\} = M \left\{ \sigma_n^2 \right\} = \sigma_n^2 \quad (2.2)$$

i.e., the power spectrum is invariant with respect to rotations. This follows from the plain fact that the integral $\int_{\sigma} f^2(\theta, \lambda) d\sigma$ is invariant

over rotations, and from Parseval's identity (1.12); (2.2) implies that the isotropic covariance function (1.10) is likewise invariant.

$$(B) \quad M \left\{ \bar{Y}_{n, m}^\alpha(p) \bar{Y}_{p, q}^\beta(q) \right\} = 0 \text{ if } \begin{cases} \alpha \neq \beta \\ n \neq p \\ m \neq q \end{cases} \text{ for all } P \text{ and } Q \quad (2.3)$$

i.e., the orthogonality properties of spherical harmonics with respect to integration on the sphere are also true with respect to averaging over rotations.

As a consequence of (2.2) and (2.3) above, the following relationships are also true:

$$M \left\{ \bar{C}_{n, m}^2 \right\} = M \left\{ \bar{S}_{n, m}^2 \right\} = \frac{\sigma_n^2}{2n+1} \quad (2.4-a)$$

$$M \left\{ \bar{C}_{n, m}^\alpha \bar{C}_{p, q}^\beta \right\} = 0 \text{ if } \begin{cases} \alpha \neq \beta \\ n \neq p \\ m \neq q \end{cases} \quad (2.4-b)$$

$$M \left\{ \bar{C}_{n, m}^\alpha \right\} = 0 \text{ if } n \neq 0 \quad (2.5)$$

$$M \left\{ \bar{C}_{n, m}^\alpha f(\theta, \lambda) \right\} = \frac{\sigma_n^2}{2n+1} \bar{Y}_{n, m}^\alpha(\theta, \lambda) \quad (2.6)$$

$$M \left\{ \bar{C}_{n, m}^\alpha \bar{f}_{1, j} \right\} = \frac{\sigma_n^2}{(2n+1)\Delta_{1, j}} \int_{\sigma_{1, j}} \bar{Y}_{n, m}^\alpha(\theta, \lambda) d\sigma \quad (2.7)$$

For the derivation of (2.6) the reader can see (Rummel, 1976), and (Sjoberg, 1978) for (2.7). As for (2.4a-b) and (2.5), the proof is given now.

According to (1.4):

$$\begin{aligned} 16\pi^2 M\{(\bar{C}_{nn}^\alpha)^2\} &= M\left\{\int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) f(\theta, \lambda) d\sigma \int_{\sigma'} \bar{Y}_{nn}^\alpha(\theta', \lambda') f(\theta', \lambda') d\sigma'\right\} = \\ &= M\left\{\int_{\sigma} \int_{\sigma'} \bar{Y}_{nn}^\alpha(\theta, \lambda) f(\theta, \lambda) \cdot f(\theta', \lambda') \bar{Y}_{nn}^\alpha(\theta', \lambda') d\sigma d\sigma'\right\} = \\ &= \int_{\sigma} \int_{\sigma'} \bar{Y}_{nn}^\alpha(\theta, \lambda) \bar{Y}_{nn}^\alpha(\theta', \lambda') M\{f(\theta, \lambda) f(\theta', \lambda')\} d\sigma d\sigma' = \\ &= \int_{\sigma} \int_{\sigma'} \bar{Y}_{nn}^\alpha(\theta, \lambda) \bar{Y}_{nn}^\alpha(\theta', \lambda') \text{cov}(f(P), f(Q)) d\sigma d\sigma' \end{aligned}$$

where $P = (\theta, \lambda)$ and $Q = (\theta', \lambda')$. According to (1.10):

$$\begin{aligned} 16\pi^2 M\{(\bar{C}_{nn}^\alpha)^2\} &= \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) d\sigma \int_{\sigma'} \bar{Y}_{nn}^\alpha(\theta', \lambda') \sum_{n=0}^{\infty} \sigma_n^2 P_n(\cos \psi_{pq}) d\sigma' = \\ &= 4\pi \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) \bar{Y}_{nn}^\alpha(\theta, \lambda) \frac{\sigma_n^2}{2n+1} d\sigma = \frac{\sigma_n^2}{2n+1} 16\pi^2 \text{ because of (1.14)} \end{aligned}$$

Similarly,

$$M\{\bar{C}_{nn}^\alpha \bar{C}_{pq}^\beta\} = \frac{\sigma_n^2}{2n+1} \cdot \frac{1}{4\pi} \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) \bar{Y}_{pq}^\beta(\theta, \lambda) d\sigma = 0 \text{ if } \begin{cases} \alpha \neq \beta \\ n \neq p \\ m \neq q \end{cases}$$

Finally, recalling that $Y_{00}^0(\theta, \lambda) = P_{00}(\cos \theta) = 1$ for all $-\pi \leq \theta \leq \pi$ and all λ ,

$$\begin{aligned} M\{\bar{C}_{nn}^\alpha\} &= \frac{1}{4\pi} \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) M\{f(\theta, \lambda)\} d\sigma = \frac{M\{f(\theta, \lambda)\}}{4\pi} \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) d\sigma \\ &= \frac{M\{f(\theta, \lambda)\}}{4\pi} \int_{\sigma} \bar{Y}_{nn}^\alpha(\theta, \lambda) Y_{00}^0(\theta, \lambda) d\sigma = 0 \quad \text{if } n \neq 0, \end{aligned}$$

which completes the proof.

2.2 Some Additional Notation

So far, data points on the sphere have been identified by the subscripts i and j . Alternatively, they could be arranged according to a single subscript $k = 2Ni + j$ (where N is the number of parallels in a grid and $2N$ the number of meridians), so the points in the "0" row, ordered by increasing j , are followed by those in the "1" row, in the same order, etc., the last element in the " $N - 1$ " row closing the sequence. Based on this convention, the set of all

values of $f(\theta_i, \lambda_j)$ (or \bar{f}_{ij}) can be arranged in N_p - vector form according to k :

$$\underline{z} = [z_0 z_1 \dots z_k \dots z_{N_p-1}]^T \quad (2.8)$$

where $z_k = f(\theta_i, \lambda_j)$ or $z_k = \bar{f}_{ij}$ with $k = 2Ni + j$

and where N_p is the number of data points, or $2N^2$ for equal angular grids. In a similar way, the coefficients \bar{C}_{nn}^α can be ordered according to a single subscript $p = n^2 + \alpha n + m + 1$ (with the understanding that the meaningless \bar{S}_{n0} are not included) defining the following N_c - vector:

$$\underline{c} = [c_0 c_1 \dots c_p \dots c_{N_c-1}]^T \quad (2.9)$$

$$c_p = \bar{C}_{nn}^\alpha \quad p = n^2 + \alpha n + m + 1$$

where $N_c = (N_{\max} + 1)^2$. Using this notation, expression (1.27) for point data quadratures can be written

$$\hat{C}_{nn}^\alpha = \underline{f}_{nn}^{\alpha T} \underline{z} \quad (2.10)$$

where the estimator vector $\underline{f}_{nn}^\alpha$, of dimension N_p , has elements of the form

$$f_{nn}^\alpha = x_i^n \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda$$

under the convention given above relating k , i and j .

Grouping all the estimates \hat{C}_{nn}^α in a vector $\hat{\underline{c}}$ ordered in the same way as \underline{c} , the relationship between the \hat{C}_{nn}^α and \underline{z} can be written, in matrix form,

$$\hat{\underline{c}} = F \underline{z} \quad (2.11)$$

where F is the estimator matrix implied by (2.10). It is a $N_p \times N_c$ matrix (where N_p is the number of data points in the grid), each row being formed by the coefficients of the quadrature formula for the corresponding \bar{C}_{nn}^α . Such row is also the transpose of the estimator vector of this \bar{C}_{nn}^α , designated $\underline{f}_{nn}^\alpha$ in (2.10).

In the same way as the covariance function between scalars, the covariance between vector functions can be defined in terms of $M\{\}$:

$$M\{\underline{z} \underline{z}^T\} = C_{zz} \quad (2.12)$$

where C_{zz} is the covariance matrix of \underline{z} , of dimension $N_p \times N_p$. This matrix is a function of the relative positions of the points in the grid on which \underline{z} has been determined, in the same way as the scalar covariance depends only on the distance between two points. The elements of C_{zz} are

$$c_{zz}^{rs} = M \{z_r, z_s\} = M \{f(P_r) f(P_s)\},$$

i.e., the values of the scalar covariance corresponding to pairs of points in the grid.

In the same way

$$M \{ \underline{c} \underline{c}^T \} = C \quad (2.13)$$

is a $N_0 \times N_0$ diagonal matrix according to (2.4a-b). C is the covariance matrix of the coefficients. Similarly, the covariance between \underline{c} and \underline{z} is

$$M \{ \underline{c} \underline{z}^T \} = C_{0z} = [M \{ \underline{z} \underline{c}^T \}]^T = C_{z0}^T \quad (2.14)$$

where C_{0z} is a $N_0 \times N_p$ matrix, the elements of which are

$$c_{0z}^{pk} = M \{ c_p, z_k \} = M \{ \bar{C}_{0z}^\alpha f(\theta_i, \lambda_j) \}$$

where the right hand side is given by (2.6).

Finally, when estimating the \bar{C}_{0z}^α , not from samples of $f(\theta, \lambda)$, but from measurements corrupted by noise

$$m(\theta_i, \lambda_j) = f(\theta_i, \lambda_j) + n_{ij} \quad (2.15)$$

the measurement errors can be grouped in a N_p -vector \underline{n} with the same ordering as \underline{z} , and the sum of both will be, then, the N_p -vector of observed values

$$\underline{m} = \underline{z} + \underline{n} \quad (2.16)$$

The measurement errors are values that occur in time, as successive observations are carried out: they constitute a time series. The average operator appropriate to them is the usual statistical expectation operator $E \{ \}$. The measurements are supposed to be unbiased, so $E \{ n_k \} = 0$ for all k . The covariance implied by this operator is the usual statistical covariance: $E \{ n_k^2 \} = \sigma_{kk}^2$, and $E \{ n_k n_r \} = \sigma_{kr}^2$. This can be generalized for the noise vector \underline{n} :

$$E \{ \underline{n} \underline{n}^T \} = D \quad (2.17)$$

where D is a $N_p \times N_p$ matrix of elements

$$d_{kr} = E \{ n_k n_r \} = \sigma_{kr}^2 \quad (2.18)$$

Both C_{zz} and D have in common a very important property:

$$\left. \begin{array}{l} \underline{x}^T C_{zz} \underline{x} \geq 0 \\ \underline{x}^T D \underline{x} \geq 0 \end{array} \right\} \text{ if } \underline{x}^T \underline{x} \neq 0, \underline{x} \text{ any } N_p \text{ vector,}$$

i.e., they are always positive matrices, moreover, in all the cases considered here, at least D is positive definite:

$$\underline{x}^T D \underline{x} > 0 \quad \text{for all } \underline{x}$$

Positiveness can be inferred readily from the definitions of C_{zz} and D :

$$\text{i.e.,} \quad \underline{x}^T D \underline{x} = \underline{x}^T E \{ \underline{n} \underline{n}^T \} \underline{x} = E \{ \underline{x}^T \underline{n} \underline{n}^T \underline{x} \} = E \{ h^2 \} \geq 0$$

(where $h = \underline{x}^T \underline{n}$), and similarly for $\underline{x}^T C_{zz} \underline{x}$ (with $M \{ \cdot \}$).

2.3 Estimation Errors, Sampling Errors, and Propagated Noise

A linear estimator is of the general form

$$\underline{s} = F \underline{m}$$

where \underline{m} is the vector of measurements defined by (2.16), and \underline{s} is the vector of estimates, made up in our case of the \hat{C}_{zz}^α . According to (2.11) and (2.16)

$$\underline{s} \equiv \underline{\hat{c}} = F (\underline{z} + \underline{n})$$

In general, the estimates will not be exactly equal to that which is estimated, the difference being the estimation error. In matrix notation

$$\underline{e} = \underline{c} - \underline{\hat{c}} = (\underline{c} - F \underline{z}) - (F \underline{n}) \quad (2.19)$$

\underline{e} being the estimation error vector. The two terms in the expression above can be defined as the components of this error:

$$\underline{e}_s = \underline{c} - F \underline{z}$$

which is the estimation error in the case of noiseless (perfect) data; and

$$\underline{e}_\eta = F \underline{n}$$

which is the error due to the noise, or propagated noise.

The error $e_{s,p}$ may be due to a number of reasons.¹ If it is zero for

¹Using the relationship $p = n^2 + \alpha n + m + 1$ of paragraph (2.2), $e_{s,p}$ stands for the sampling error in \hat{C}_{zz}^α , and $e_{\eta,p}$ for the propagated noise.

some estimator, then its presence in other estimators could be blamed on them being somewhat inadequate. For instance, if the estimator was chosen by taking the elements of F from a set of random numbers, then the estimation error is likely to be always high, as the estimator has nothing to do with the actual problem. In particular, the addition of extra measurements to the vector \underline{m} is not going to bring any general improvement on the estimates. On the other hand, if attention is paid to the nature of the problem when selecting F , one would expect the error to decrease as more data is introduced. If, as the number N_s of samples in \underline{m} tends to infinity, $e_{s,p}$ tends to zero, one could say that the error is due to the incomplete sampling of the signal $f(\theta, \lambda)$, and call it the sampling error. This is precisely the case with any of the quadrature formulas to be studied here, all of which can be written formally as linear estimators according to (2.11), and for all of which the error $e_{s,p}$ vanishes as the number of samples tends to infinity, because the sums become identical with the integrals defined by (1.4). In this sense it is quite suitable to call $e_{s,p}$ the sampling error, as in paragraph (1.3).

2.4 The Quadratic Error Measure

The overall error measure will be defined here as the sum of two quadratic terms: one for the propagated noise, the other for the sampling error.

(a) Propagated Noise Measure

This measure is the same as in least squares adjustment, i.e., the variance of the error defined in terms of the usual statistical expectation operator

$$\sigma_{\eta}^2 = E \left\{ \left(\underline{f}_{ns}^{\alpha} \underline{n} \right)^2 \right\} \quad (2.20)$$

according to (2.10). This variance represents the scatter in the value of \hat{C}_{ns}^{α} due to the uncertainty in the values of the data. In matrix form

$$E_{\eta} = E \left\{ \underline{e}_{\eta} \underline{e}_{\eta}^T \right\} = E \left\{ F \underline{n} \underline{n}^T F^T \right\} = F E \left\{ \underline{n} \underline{n}^T \right\} F^T = F D F^T \quad (2.21)$$

where E_{η} is a $N_s \times N_s$ matrix, while D was already presented in paragraph (2.2).

In the special case where the measurement errors are uncorrelated, D is diagonal, and (2.20) becomes

$$\sigma_{\eta}^2 = E \left\{ \underline{f}_{ns}^{\alpha} \underline{n} \underline{n}^T \underline{f}_{ns}^{\alpha} \right\} = \underline{f}_{ns}^{\alpha} D \underline{f}_{ns}^{\alpha} = \sum_{i=0}^{N-1} (\chi_i^{\alpha})^2 \sum_{j=0}^{2N-1} \left\{ \begin{matrix} \cos^2 \\ \sin^2 \end{matrix} \right\} m_j \Delta \lambda E \left\{ n_{ij}^2 \right\} \quad (2.22)$$

which is the usual formula for propagating the covariance of the noise.

(b) Sampling Error Measure

This measure is defined in terms of the isotropic averaging operator of paragraph (2.1)

$$\sigma_{s\ n\alpha}^2 = M \left\{ (c_p - \underline{f}_{n\alpha}^\alpha \underline{z})^2 \right\} \equiv M \left\{ (\bar{C}_{n\alpha}^\alpha - \underline{f}_{n\alpha}^\alpha \underline{z})^2 \right\} \quad (2.23)$$

or, in matrix form

$$E_s = M \left\{ \underline{e}_s \underline{e}_s^T \right\} = M \left\{ (\underline{c} - F \underline{z})(\underline{c} - F \underline{z})^T \right\} \quad (2.24)$$

$$= C - 2 C_{sz} F^T + F C_{zz} F^T$$

where E_s is a $N_s \times N_s$ matrix, and C , C_{sz} , C_{zz} were introduced in paragraph (2.2).

(c) Total Error Measure

The total measure is the sum of (a) and (b)

$$\sigma_{n\alpha}^2 = \sigma_{\eta\ n\alpha}^2 + \sigma_{s\ n\alpha}^2 \quad (2.25)$$

or, in matrix form,

$$E_T = E_s + E_\eta = C - 2C_{sz} F^T + F C_{zz} F^T + F D F^T = \quad (2.26)$$

$$C - 2C_{sz} F^T + F (C_{zz} + D) F^T$$

where E_T is the $N_p \times N_p$ error matrix associated with F and with the covariances that define C , C_{sz} and $(C_{zz} + D)$. Expression (2.26) is a special case of the formula for " E_{ss} " in least squares collocation (for instance, Moritz (1978), Ch. 3, eqn, (3.20)); moreover, it belongs to a family of formulas also found in the minimum variance estimation and filtering of time series and of processes sampled on the euclidean plane.

The total measure has been chosen simply as the sum of $\sigma_{s\ n\alpha}^2 + \sigma_{\eta\ n\alpha}^2$ by making the basic assumption that the sampling error and the propagated noise are due to completely independent causes. The first depends on the values $f(\theta_i, \lambda_j)$, while the second depends on the measurement errors of instruments that, at least ideally, operate with accuracies unaffected by the quantities measured, or in such way that any interactions can be eliminated by simple corrections.

The columns of F are defined by the quadrature formula used, and such formulas either satisfy, or tend to satisfy, orthogonality conditions (paragraph (1.3(c))). For this reason, provided that C_{zz} and D belong to the type to be described in paragraph (2.9), matrices E_η , E_s , and, thus, E_T , are either diagonal or diagonal dominant, and in the latter case tend to become diagonal as

the sampling intervals decrease, or $N_p \rightarrow \infty$. For this reason the correlations among the errors for individual coefficients are, or "tend to be", very small.

The diagonal elements of the error matrices E_s , E_η and E_τ are the variances of the errors in the respective coefficients, as defined by (2.20), (2.23), and (2.25), respectively.

2.5 The Meaning of the Error Measure

The treatment of the propagated noise is the same as in least squares adjustment, so this part of the error measure should be easily understood. The sampling error measure, on the other hand, is a geometrical measure: $M\{\}$ belongs, as a concept, in the field of integral geometry, or the study of "geometric probabilities". This is a branch of mathematics closely related to integration and to measure theory, and also to statistical mechanics. In geodesy, this type of idea is relatively new (Kaula, 1959), Moritz (1965), but it has been used already extensively enough to show its considerable worth.

From expressions (1.10) and (1.11), the covariance and the power spectrum are functions of each other. Since either of them, and the sampling grid, define matrices C , $C_{\alpha\alpha}$ and C_{zz} in expression (2.24), it follows that a statement on $\sigma_{\alpha\alpha}^2$ is, somehow, also a statement on the performance of F for all the functions that have the same power spectrum that determines the diagonal elements of C . To put this more precisely, consider a function $f_1(\theta, \lambda)$ having the given power spectrum. If $\bar{C}_{\alpha\alpha}^\alpha$ were estimated for f_1 and also, at least ideally, for all its rotations, then the mean square of the sampling error $e_{\alpha p}$ in $\bar{C}_{\alpha\alpha}^\alpha$ for all this functions would be, by definition of $M\{\}$, the measure $\sigma_{\alpha\alpha}^2$. If a second function f_2 (perhaps not a rotation of f_1) and all its rotations were then analysed in the same way, the average of $e_{\alpha p}^2$ for all these functions would be, once more, $\sigma_{\alpha\alpha}^2$, as long as f_2 has the same spectrum as f_1 . Moreover, the average of $e_{\alpha p}^2$ for f_1 , f_2 , and their rotations put together, would also be $\sigma_{\alpha\alpha}^2$. In fact, if we had a finite set of functions f_1, f_2, \dots, f_n , with arbitrary n , all with the same power spectrum (or covariance), then $e_{\alpha p}^2$ would average $\sigma_{\alpha\alpha}^2$ for all the f_i and their rotations.

It appears, from the preceeding discussion, that one could take a simple step and say " $\sigma_{\alpha\alpha}^2$ is the mean of the sampling error squared of the estimator $\hat{C}_{\alpha\alpha}^\alpha = \underline{f}_{\alpha\alpha}^\alpha \underline{z}$, over all possible functions with the given power spectrum." Unfortunately, as mentioned in the introduction, the sphere is a rather wicked surface. There is a theorem by Lauritzen (1973) that states the impossibility of having the same average $\sigma_{\alpha\alpha}^2$ for all functions as for every function, when the distribution of the ensemble happens to be gaussian. Moritz (1978) has endeavoured to show that this is no problem if the ensemble of functions

is not gaussian, but using his conclusions here would force the introduction of a rather strange requirement of "non-gaussness" on the ensemble of the signals analysed that is best left out, if possible.

Perhaps there is a way out in going back to the idea of a finite set of functions f_i , where the problem does not exist, by saying:

"the error measure $\sigma_{n\alpha}^2$, for a certain estimator and a certain type of signal power spectrum, is the average of the square of the sampling error in $\hat{C}_{n\alpha}^{\alpha}$ for all functions with the given power spectrum EVER TO BE ANALYSED with that estimator, and for all their rotations".

After all, accuracy is what geodesists are always interested in, not perfection.

2.6 Simple Formulas for Area Means

The numerical studies of section 3 concentrate in area mean type formulas, because area means are preferred for collating information, particularly on a global basis, at present. The formulas to be studied here and in that section can be divided into "simple" and "optimal". The name "simple" is given here to expressions of the type

$$\hat{C}_{n\alpha}^{\alpha} = \mu_n \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} T_{ij} \int_{\sigma_{ij}} \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda) d\sigma \quad (2.27)$$

where μ_n is a scale factor affecting the n th harmonic as a whole. Expressions of this type have been developed more or less intuitively, along the lines of the following reasoning:

If the signal were constant on each block, it will equal its mean value there, and the coefficients of such a function would be precisely

$$\bar{C}_{n\alpha}^{\alpha} = \frac{1}{4\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} T_{ij} \int_{\sigma_{ij}} \bar{Y}_{n\alpha}^{\alpha}(\theta, \lambda) d\sigma \quad (2.28)$$

according to (1.4). In general, most signals are not equal to their mean value over whole blocks, so the expression would not be exact. In most cases, the signal would have fluctuations in each block, and it would be less smooth than a function that is constant over each block, so using the formula above with T_{ij} as data may result in the $\tilde{C}_{n\alpha}^{\alpha}$ of a smoothed function. As a refinement, one could try to de-smooth the $\tilde{C}_{n\alpha}^{\alpha}$. If the blocks were circular, the relationship between "true" and "smooth" $\bar{C}_{n\alpha}^{\alpha}$ would be

$$\bar{C}_{nn}^{\alpha} = \frac{1}{\beta_n} \bar{C}_{nn}^{\alpha} \quad (2.29)$$

where β_n is known as the Pellinen smoothing factor of degree n . The relationship between β_n and the radius of the circular blocks is given in paragraph (4.3). For small blocks, experience shows that there is little difference between the area means of geodetic data on circular or on square blocks, so the error is small if one assumes that they are the same; in such case the modified expression

$$\hat{C}_{nn}^{\alpha} = \frac{1}{4\pi\beta_n} \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{f}_{ij} \int_{\sigma_{ij}} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma = \bar{C}_{nn}^{\alpha} \quad (2.30)$$

could be used; in practice, this is only an approximation, though a good one, as showed by Katsambalos (1979), who tested this expression extensively.

In addition to (2.28) and (2.30), Lowes (1978) has proposed using

$$\hat{C}_{nn}^{\alpha} = \frac{1}{4\pi\beta_n^2} \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{f}_{ij} \int_{\sigma_{ij}} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma \quad (2.31)$$

to estimate the harmonic coefficients. All these expressions have the property that, because $\frac{1}{\beta_n} \rightarrow 1$, and $\sum_{i,j} \bar{f}_{ij} \int_{\sigma_{ij}} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma \rightarrow \int_{\sigma_{ij}} f(\theta, \lambda) \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma$ as $\Delta_{ij} \rightarrow 0$ (or $N_p \rightarrow \infty$), it is true that the error $e_{sp} = \bar{C}_{nn}^{\alpha} - \hat{C}_{nn}^{\alpha} \rightarrow 0$ with $N_p \rightarrow \infty$; in other words: e_{sp} is properly called a sampling error in the sense given to this term in paragraph (2.3).

Comparing (2.28), (2.30), and (2.31) it is easy to see that they all belong to a class of expressions of the form (2.27), with $\mu_n = \frac{1}{4\pi}$, $\mu_n = \frac{1}{4\pi\beta_n}$, and $\mu_n = \frac{1}{4\pi\beta_n^2}$, respectively.

The scaling factor μ_n can also be regarded as a de-smoothing factor, if one wishes to retain the intuitive meaning of these formulas. In the notation of paragraph (2.2), these expressions can be written, according to (2.27), as

$$\hat{C}_{nn}^{\alpha} = \mu_n (\underline{h}_{nn}^{\alpha})^T \underline{z} \quad (2.32)$$

with $\mu_n \underline{h}_{nn}^{\alpha} = \underline{f}_{nn}^{\alpha}$.

Replacing (2.32) in the definition of the sampling error measure, (2.23), and adding with respect to m and α to obtain the total error in the n th harmonic:

$$\begin{aligned} \sum_{\alpha=0}^1 \sum_{n=0}^{\infty} \sigma_{s,nn}^2 &= \sigma_n^2 - \left[2 \sum_{\alpha=0}^1 \sum_{n=0}^{\infty} \underline{c}_{nn}^T \alpha_z \underline{h}_{nn}^{\alpha} \right] \mu_n + \\ &\left[\sum_{\alpha=0}^1 \sum_{n=0}^{\infty} (\underline{h}_{nn}^{\alpha})^T (C_{zz}) \underline{h}_{nn}^{\alpha} \right] \mu_n^2 \end{aligned} \quad (2.33)$$

(where $\underline{c}_{n\alpha, z}^T$ is a row of C_{zz}). This is the sum of certain diagonal elements of E_s , according to (2.24), when the estimator has the form (2.27). Clearly, (2.33) is a quadratic function of the scalar μ_n , and as such it can have either a maximum or a minimum. If C_{zz} is positive definite, it must be a minimum. Finding the corresponding value of μ_n is the same as finding the formula of type (2.27) that has the smallest sampling error per harmonic for signals with the covariance (power spectrum) specified by C_{zz} . In addition to the sampling error, the measure of the propagated noise can be added to obtain

$$\sum_{\alpha=0}^1 \sum_{n=0}^n \sigma_{nn}^{2\alpha} = \sigma_n^2 - \left[\sum_{\alpha=0}^1 \sum_{n=0}^n 2 \underline{c}_{n\alpha, z}^T \underline{h}_{n\alpha}^\alpha \right] \mu_n + \left[\sum_{\alpha=0}^1 \sum_{n=0}^n (\underline{h}_{n\alpha}^\alpha)^T (C_{zz} + D) \underline{h}_{n\alpha}^\alpha \right] \mu_n^2 \quad (2.34)$$

This is also quadratic and has a minimum, and finding the optimum μ_n is the subject of the next paragraph.

2.7 Optimum de-Smoothing Factors

The coefficients of μ_n^2 and μ_n in (2.33) and (2.34) are both real scalars, and so is the independent term σ_n^2 . The expressions represent parabolas, and because both C_{zz} and D are positive, if the further (and likely) assumption is made that they are also definite, then $\frac{\partial^2}{\partial \mu_n^2} \sum_{\alpha} \sum_{n} \sigma_{nn}^{2\alpha} > 0$,

and the parabola has a minimum where μ_n satisfies the condition

$$\frac{1}{2} \frac{\partial}{\partial \mu_n} \sum_{\alpha} \sum_{n} \sigma_{nn}^{2\alpha} = - \sum_{\alpha} \sum_{n} \underline{c}_{n\alpha, z}^T \underline{h}_{n\alpha}^\alpha + \left[\sum_{\alpha} \sum_{n} (\underline{h}_{n\alpha}^\alpha)^T C_{zz} \underline{h}_{n\alpha}^\alpha \right] \mu_n \quad \text{for the sampling error (2.33)}$$

$$\text{i.e., at } \hat{\mu}_n = \frac{\sum_{\alpha} \sum_{n} \underline{c}_{n\alpha, z}^T \underline{h}_{n\alpha}^\alpha}{\sum_{\alpha} \sum_{n} (\underline{h}_{n\alpha}^\alpha)^T C_{zz} \underline{h}_{n\alpha}^\alpha} \quad (2.35)$$

$$\text{or at } \mu_n^* = \frac{\sum_{\alpha} \sum_{n} \underline{c}_{n\alpha, z}^T \underline{h}_{n\alpha}^\alpha}{\sum_{\alpha} \sum_{n} (\underline{h}_{n\alpha}^\alpha)^T (C_{zz} + D) \underline{h}_{n\alpha}^\alpha} \quad (2.36)$$

for the total error (2.34).

Expressions (2.28), (2.30), (2.31), and (2.36) will be studied further, by means of computed examples, in section 3.

2.8 Least Squares Collocation

In the notation of Paragraph (2.2), optimizing μ_n is the same as obtaining the optimal vectors $\underline{f}_{nn}^\alpha$ of the form

$$\underline{f}_{nn}^\alpha = \mu_n \underline{h}_{nn}^\alpha$$

for the estimator

$$\hat{\underline{c}} = F \underline{m}$$

where $\underline{m} = \underline{z} + \underline{n}$, and the $(\underline{f}_{nn}^\alpha)^\top$ are the rows of F . If no restriction is placed on the form the rows of F can take, then a reasoning similar to that in the preceding paragraph leads to the best possible linear estimator for the $\hat{\underline{c}}_{nn}^\alpha$.

Considering the total measure of error for $0 \leq n \leq N$:

$$\begin{aligned} \sigma_\epsilon^2 &= \sum_{\alpha=0}^I \sum_{n=0}^{N-1} \sum_{m=0}^n \sigma_{nm}^2 \alpha = \sum_{n=0}^{N-1} \sigma_n^2 - 2 \sum_{\alpha=0}^I \sum_{m=0}^{N-1} \sum_{n=0}^m \underline{c}_{nm}^\top \alpha_z \underline{f}_{nn}^\alpha + \\ &\sum_{\alpha=0}^I \sum_{n=0}^{N-1} \sum_{m=0}^n (\underline{f}_{nn}^\alpha)^\top (C_{zz} + D) \underline{f}_{mm}^\alpha \end{aligned} \quad (2.37)$$

it is not difficult to see that, because all σ_{nm}^2 are non-negative, finding the F that minimizes their sum is the same as finding the F that minimizes them individually. The sum of the mean squared errors of all coefficients is the trace of the error matrix E_T of (2.26):

$$\sigma_\epsilon^2 = \sum_{\alpha=0}^I \sum_{n=0}^{N-1} \sum_{m=0}^n \sigma_{nm}^2 \alpha = \text{tr}[E_T]$$

To obtain the condition for a minimum, one must differentiate (2.37) so, according to (2.26),

$$\frac{1}{2} \frac{\partial \text{tr}}{\partial F} [E_T] = -C_{0z}^\top + (C_{zz} + D) F^\top = 0 \quad (2.38)$$

as found using well-known matrix analysis formulas. From this follows that

$$F = C_{0z} (C_{zz} + D)^{-1} \quad (2.39)$$

is the F that minimizes (2.37), provided that $(C_{zz} + D)$ is positive definite. As already explained, both matrices are always positive and their sum is usually definite. The expression for the optimal estimator for $\hat{\underline{c}}_{nn}^\alpha$ is

$$\hat{\underline{c}}_{nn}^\alpha = (\underline{f}_{nn}^\alpha)^\top \underline{m} = \underline{c}_{nm}^\top \alpha_z (C_{zz} + D)^{-1} \underline{m} \quad (2.40)$$

where $(\hat{f}_{\alpha\alpha}^*)^T$ is the row of the optimal estimator matrix F corresponding to $\hat{C}_{\alpha\alpha}$.

The use of expression (2.40) is, in brief, least squares collocation applied to spherical harmonic analysis.

When the optimal F is used, the error matrix becomes, according to (2.26) and (2.39),

$$E_r = C - C_{\alpha z} (C_{zz} + D)^{-1} C_{\alpha z}^T \quad (2.41)$$

and the total error measure is the trace of this matrix: by definition, the smallest for all possible F .

Clearly, whether one is interested in estimating coefficients or in determining the likely accuracy of such estimates, using expressions (2.40) or (2.41) require a knowledge of either $(C_{zz} + D)^{-1}$ (inversion) or, at least, of $C_{\alpha z} (C_{zz} + D)^{-1}$ (solution). Because the matrix $(C_{zz} + D)$ has dimension $N_p \times N_p$, obtaining either requires, by usual linear algebra methods, $O(N_p^3)$ (or $O(N_p^6)$) operations. In the case of a $1^\circ \times 1^\circ$ grid, $N_p = 64800$, so, at some 200000 products and sums (double precision) per second, a modern computer like the one at OSU would need about one century to obtain all $\hat{C}_{\alpha\alpha}$ to degree and order 180 from data on such a grid. Fortunately, as explained in paragraph (2.9), if the covariance functions of signal and noise both satisfy certain conditions, and if $\Delta\lambda$ is constant for the whole grid, then both C_{zz} and D (consequently their sum) can be inverted in much fewer operations than by conventional methods, because they possess a particularly strong structure. Moreover, the optimal estimator $\hat{C}_{\alpha\alpha} = (\hat{f}_{\alpha\alpha}^*)^T \underline{m}$ turns out to be of the form (1.24) or (1.7), depending on the kind of data \underline{m} , so, under rather general conditions, the optimal estimator of $\hat{C}_{\alpha\alpha}$ is also the best quadratures type formula for point data or for area means, as the case may be.

The conditions mentioned above are satisfied, for instance, when both the geometrical covariance $\text{cov}(f(P), f(Q))$ and the stochastic covariance $E\{n_1, n_2\}$ ($P \equiv (\theta_1, \lambda_1)$, $Q \equiv (\theta_2, \lambda_2)$) are isotropic, i.e., functions only of the separation between the points P and Q . By definition of $M\{\}$, the geometrical covariance obtained using this operator is isotropic, so C_{zz} has the desired structure. A common assumption regarding good instruments is that the n_{1j} are uncorrelated, so D is diagonal. If the errors are stationary, so their variances are constant, or at least constant along parallels, then matrix D has the required structure, and inverting $C_{zz} + D$ can be greatly expedited. In practice, however, this is not likely to be the case, as the number and quality of measurements will vary from region to region, resulting in different σ_1^2 , both globally and along parallels. As a result, the best linear estimator in terms of the chosen error measure will not have the

quadratures form, and it will be very difficult to compute when the number of data values is very large. Nevertheless, as shown in section 3, quadrature formulas can give reasonable estimates of the \bar{C}_{nn}^α with noisy data sets where the noise is uneven, so it would be interesting to get the best quadrature formula for a particular combination of signal and noise, provided that such a formula can be obtained without undue effort.

2.9 The Best Quadrature Formula for non-Uniform, Uncorrelated Noise

If the variance of the noise fluctuates along parallels, matrix D , though diagonal, is such that the minimization of the error measure (2.37)

$$\sigma_\epsilon^2 = \text{tr} [C - 2C_{sz} F^T + F(C_{zz} + D) F^T] = \hat{\Phi}(D) \quad (2.42)$$

(see also (2.26))

can be very difficult with large N_p , and the optimal estimator is not of the quadrature type. Introducing a "modified noise matrix" L , also diagonal and where the diagonal elements are

$$l_{kk} = \frac{1}{2N} \sum_{j=0}^{2N-1} \sigma_{1j}^2 \quad (\text{with } k = 2Nl + j) \quad (2.43)$$

the following modified error measure $\hat{\Phi}(L)$ can be defined:

$$\hat{\Phi}(L) = \text{tr} [C - 2C_{sz} F^T + F(C_{zz} + L) F^T] \quad (2.44)$$

The optimal estimator for this measure is easy to obtain, and is of the quadratures type.

The parts of $\hat{\Phi}(D)$ and $\hat{\Phi}(L)$ that measure the sampling errors are identical, so any difference between the overall measures must come from the "noise propagation" parts $\text{tr} [F D F^T]$ and $\text{tr} [F L F^T]$. If the estimator (not necessarily optimal) happens to be of the quadratures type, i.e., for point data:

$$\hat{C}_{nn}^\alpha = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} x_i^{ns} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda [f(\theta_i, \lambda_j) + n_{1j}] \quad (2.45)$$

then the propagated noise is, assuming the n_{1j} to be uncorrelated,

$$\sigma_{\eta nn}^{2\alpha} = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} (x_i^{ns})^2 \begin{Bmatrix} \cos^2 \\ \sin^2 \end{Bmatrix} m_j \Delta \lambda E\{n_{1j}^2\}$$

so

$$\hat{\Phi}_{\eta}(D) = \text{tr} [F D F^T] = \sum_{\alpha=0}^1 \sum_{n=0}^{N-1} \sum_{m=0}^N \alpha_{\eta n}^{2\alpha} = \sum_{\alpha=0}^1 \sum_{n=0}^{N-1} \sum_{i=0}^{N-1} (x_i^{ns})^2 \sum_{j=0}^{2N-1} \sigma_{1j}^2 \quad (2.46)$$

The "modified noise" on the other hand, is, according to (2.43) and (2.44):

$$\begin{aligned} \hat{\sigma}_\eta^2(L) &= \text{tr} [F L F^T] = \sum_{n=0}^{N-1} \sum_{m=0}^2 \sum_{l=0}^{N-1} (x_l^{n,m})^2 \sum_{j=0}^{2N-1} (\cos^2 m j \Delta\lambda + \sin^2 m j \Delta\lambda) \cdot \\ &\cdot \frac{1}{2N} \sum_{j=0}^{2N-1} \sigma_{1,j}^2 = \sum_{n=0}^{N-1} \sum_{m=0}^2 \sum_{l=0}^{N-1} (x_l^{n,m})^2 \sum_{j=0}^{2N-1} \sigma_{1,j}^2 \end{aligned} \quad (2.47)$$

Comparing the expressions for $[F L F^T]$ and for $[F D F^T]$, it follows that they are identical, and since the "sampling" parts are also identical in (2.42) and (2.44), then

$$\text{tr} [C - 2C_{ez} F^T + F(C_{zz} + L) F^T] = \text{tr} [C - 2C_{ez} F^T + F(C_{zz} + D) F^T] \quad (2.48)$$

This means that the actual and the modified error measures must coincide if the estimator is of the quadratures type.

Replacing D with L in equation (2.38) and solving for the estimator matrix, one gets

$$F_L = C_{ez}(C_{zz} + L)^{-1} \quad (2.49)$$

where F_L is the estimator matrix that minimizes the modified error measure (2.44). Because of the way L has been defined, this estimator is of the quadratures type, so the modified and the actual error measures coincide, as just shown.

Assume that there is an estimator, different from $\hat{\underline{c}} = F_L \underline{m}$ but also of the quadratures type, the estimator matrix of which is \tilde{F} , and such that:

$$\text{tr} [C - 2C_{ez} \tilde{F}^T + \tilde{F}(C_{zz} + D) \tilde{F}^T] < \text{tr} [C - 2C_{ez} F_L^T + F_L(C_{zz} + D) F_L^T]$$

Then, according to (2.48),

$$\text{tr} [C - 2C_{ez} \tilde{F}^T + F(C_{zz} + L) \tilde{F}^T] < \text{tr} [C - 2C_{ez} F_L^T + F_L(C_{zz} + L) F_L^T] \quad (2.50)$$

which contradicts the fact that F_L minimizes the modified error measure. Therefore, (2.50) cannot be true, and F_L must be the matrix of the optimal quadratures type estimator that minimizes the actual error measure σ_ϵ^2 .

The optimal quadratures type estimator, as the name indicates, is the best of a certain kind, not the absolute best. The best estimator, when no conditions as to its form are imposed, will not be (in general) of the quadratures type, unless D happens to have the "right form" specified before, i.e., unless $D = L$.

When $D = L$, the optimal estimator and the best quadrature formula coincide. Regardless of this, the quadrature formula obtained from (2.49) is the best, so its error measure is a lower bound for those of all other quadratures formulas with the given signal and noise.

When $D \neq L$, while minimizing the sum of all error variances $\sigma_{nn}^{2\alpha}$, i.e., $\text{tr}(E_T)$, the optimal quadrature formula does not minimize each individual variance $\sigma_{nn}^{2\alpha}$. To show this, consider the propagated noise measure for C_{nn}^α when the n_{ij} are uncorrelated (to simplify the argument):

$$\sigma_{nn}^{2\alpha} = \sum_{i=0}^{N-1} (x_i^{nn})^2 \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix}^2 m_j \Delta \lambda \sigma_{ij}^2 = \sum_{i=0}^{N-1} (x_i^{nn})^2 \left[\frac{1}{2} \sum_{j=0}^{2N-1} \sigma_{ij}^2 + \frac{1}{2} \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ -\sin \end{Bmatrix} 2m_j \Delta \lambda \sigma_{ij}^2 \right]$$

The modified error measure, minimized by the formula, is

$$\sum_{i=0}^{N-1} (x_i^{nn})^2 \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos^2 \\ \sin^2 \end{Bmatrix} m_j \Delta \lambda \frac{1}{2N} \sum_{j=0}^{2N-1} \sigma_{ij}^2 = \sum_{i=0}^{N-1} (x_i^{nn})^2 \frac{1}{2} \sum_{j=0}^{2N-1} \sigma_{ij}^2$$

Clearly, both are not the same, unless

$$\sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ -\sin \end{Bmatrix} 2m_j \Delta \lambda \sigma_{ij}^2 = 0$$

which is not likely to be fulfilled for arbitrary σ_{ij}^2 . However, looking at the reasoning which leads to (2.47), one can see that the sums of the modified and the actual error measures for pairs $(\bar{C}_{nn}, \bar{S}_{nn})$, and also for the individual \bar{C}_{n0} , are already identical. From this follows that the variance of the error per degree

$$\sum_{\alpha=0}^1 \sum_{n=0}^2 \sigma_{nn}^{2\alpha} = \sigma_{\epsilon_n}^2 \quad (2.51)$$

and per average coefficient per degree:

$$\delta \epsilon_n^2 = \frac{\sigma_{\epsilon_n}^2}{2n+1} \quad (2.52)$$

are also identical to the modified measure. So, while nothing can be predicated of individual coefficients, the error for each harmonic as a whole and that for the "average coefficient" in it are going to be minimum. By Parseval's identity (1.12), if the coefficients were used to calculate, say, geoidal undulations, the mean squared error of the computed geoid, globally, would be the same as the sum of the error squared of the normalized coefficients, so individual coefficient variances are of little interest in this and similar applications, while the $\sigma_{\epsilon_n}^2$ are very important. This shows that the optimal quadratures formula when $D \neq L$ can be just as useful as when $D = L$.

The discussion in this paragraph has been centered on point value type formulas; the conclusions apply equally well to area mean type formulas, the extension of the reasoning being quite straightforward.

2.10 The Structure of the Covariance Matrix and its Consequences

The following discussion summarizes some results presented by this author in a previous report (Colombo, 1979a). In order to be able to calculate the variance of the error $\sigma_{nn}^{2\alpha}$ with expressions such as (2.26), and also to be able to obtain the optimal estimator according to collocation theory, it is necessary to create and invert the $N_p \times N_p$ matrix $(C_{zz} + D)$, which can be very

large if the number of data N_p is large. In the case of regularly sampled data this two problems can be greatly simplified if the covariances and the grid has certain symmetries. The most important of these are: (a) the sampling in longitude must be at constant intervals and along parallels (or parallel bands, i.e., rows of blocks); (b) for given i and p the covariances $\text{cov}(u(\theta_i, \lambda_j), v(\theta_p, \lambda_q))$ (or $\text{cov}(\bar{u}_{ij}, \bar{v}_{pq})$) and $E\{n_{ij}, n_{pq}\}$, must depend only on $|j-q|$. It is also very advantageous, though not essential, that the grid be symmetrical with respect to the Equator.

In what follows N_r is the number of parallels and N_l the number of meridians ($N_r = N$, $N_l = 2N$ when the grid is equal angular).

Under this set of conditions, if the data vector \underline{m} is ordered according to (2.8) and is subdivided into partitions \underline{m}_i , where

$$\underline{m}_i = [m_{i0} m_{i1} \dots m_{iN_l-1}]^T$$

includes all data values in the same parallel or row of blocks, then the matrix $(C_{zz} + D)$ can be partitioned into N_r^2 blocks C^{ip} , each of dimension $N_l \times N_l$, containing the covariances between the data along rows i and p .

Each block C^{ip} has a Toeplitz circulant structure, because its elements satisfy the relationships

$$c_{jq}^{ip} = c_{j+1, q+1}^{ip} ; c_{j0}^{ip} = c_{j-1, N_l-1}^{ip} \quad \text{when } j > 0$$

which follow from the fact that parallels are circular, and that the covariance between points in parallels i and p is a function of $|j-q|$. Moreover, the elements in the first row or column (the C^{ip} are symmetrical) also satisfy

$$c_{0q}^{ip} = c_{0, N_l-q}^{ip} \quad \text{when } q > 0$$

Therefore, the first row can be represented exactly as a sum of $\frac{1}{2}N_l + 1$ cosines:

$$c_{0q}^{ip} = \sum_{n=0}^{\frac{1}{2}N_l} a_n^{ip} \cos n \frac{2\pi}{N_l} q \quad (2.53)$$

The a_n^{ip} form the discrete Fourier transform of the sequence

$$c_{00}^{ip}, c_{01}^{ip}, \dots, c_{0, N_l-1}^{ip}$$

If

$$r_{\frac{1}{2}}^{ip} = H a_{\frac{1}{2}}^{ip} \quad (2.54)$$

$$\text{where } H = \begin{cases} N1 & \text{if } m = 0 \\ \frac{N1}{2} & \text{if } m \neq 0 \end{cases}$$

If $R(m)$ is the matrix where each "ip" element equals $r_{\frac{1}{2}}^{ip}$, then (as explained by Colombo, (op. cit.)) inverting $(C_{zz} + D)$ is equivalent to inverting the $Nr \times Nr$ matrices $R(m)$ for $m = 0, 1, \dots, \frac{1}{2}N1$. Isotropic covariances satisfy the " $|j - q|$ condition" mentioned above, so, for fixed $\Delta\lambda$, the covariance matrix always has this regular structure.

Let

$$\underline{c}_{\frac{1}{2}}^{\alpha} = \left[\begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m0\Delta\lambda, \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m\Delta\lambda, \dots, \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m(N1-1)\Delta\lambda \right]^T$$

A vector of the type

$$\underline{v}^{\alpha} = [v_0 \underline{c}_{\frac{1}{2}}^{\alpha T}, v_1 \underline{c}_{\frac{1}{2}}^{\alpha T}, \dots, v_{Nr-1} \underline{c}_{\frac{1}{2}}^{\alpha T}]^T \quad (2.55)$$

shall be called, for convenience, a vector "of frequency m".

Under the conditions described before all the eigenvectors of $(C_{zz} + D)$ are vectors of frequency m, with $m = 0, 1, \dots, \frac{1}{2}N1$. Moreover, if $\lambda_{t,\frac{1}{2}}$ ($t = 1, 2, \dots, Nr$) is one of the Nr eigenvalues of $R(m)$, and if

$$\underline{s}_{t,\frac{1}{2}} = [s_0^{t,\frac{1}{2}} \dots s_{Nr-1}^{t,\frac{1}{2}}]^T$$

is the corresponding eigenvector of $R(m)$, then $\lambda_{t,\frac{1}{2}}$ is also an eigenvalue of $(C_{zz} + D)$, and the pair

$$\underline{s}_{t,\frac{1}{2}}^{\alpha} = [s_0^{t,\frac{1}{2}} \underline{c}_{\frac{1}{2}}^{\alpha T}, \dots, s_{Nr-1}^{t,\frac{1}{2}} \underline{c}_{\frac{1}{2}}^{\alpha T}]^T$$

the two corresponding eigenvectors of $(C_{zz} + D)$. Therefore, to decompose the large covariance matrix in eigenvectors and eigenvalues is equivalent to decomposing the $\frac{1}{2}N1 + 1$ matrices $R(m)$, and this is why the latter are relevant to the inversion of $(C_{zz} + D)$: the eigenvalues of the inverse are the reciprocal of the $\lambda_{t,\frac{1}{2}}$, while its eigenvectors are the same as the $\underline{s}_{t,\frac{1}{2}}$. Further, this implies that $(C_{zz} + D)^{-1}$ has the same structure as the covariance matrix, i.e., it consists of Toeplitz-circulant blocks.

Since $(C_{zz} + D)^{-1}$ has eigenvectors of frequency m , then, if \underline{h} is a linear combination of vectors of a given frequency, $\underline{z} = (C_{zz} + D)^{-1} \underline{h}$ is also a linear combination of vectors of that frequency. In the case of point data, from expression (2.6) follows that the cross-covariances vector in (2.40) is

$$\begin{aligned} \underline{c}_{nn}^T \alpha_{z,j} &= \frac{\sigma_n^2}{2n+1} [\bar{P}_{nn}(\cos \theta_0) \underline{c}_n^{\alpha^T} \dots \bar{P}_{nn}(\cos \theta_{Nr-1}) \underline{c}_n^{\alpha^T}] \\ &= [k_0^{nn} \underline{c}_n^{\alpha^T} \dots k_{Nr-1}^{nn} \underline{c}_n^{\alpha^T}] \end{aligned} \quad (2.56)$$

Define

$$\underline{k}^{nn} = [k_0^{nn} \dots k_{Nr-1}^{nn}]^T$$

and

$$\underline{\chi}^{nn} = [\chi_0^{nn} \dots \chi_{Nr-1}^{nn}]^T$$

Then $\underline{f}_{nn}^{\alpha} = (C_{zz} + D)^{-1} \underline{c}_{nn}^{\alpha} \alpha_{z,j}$ must be of the form

$$\underline{f}_{nn}^{\alpha} = [\chi_0^{nn} \underline{c}_n^{\alpha^T} \dots \chi_{Nr-1}^{nn} \underline{c}_n^{\alpha^T}]^T \quad (2.57)$$

where, according to Colombo (ibid),

$$\underline{\chi}^{nn} = R(m)^{-1} \underline{k}^{nn} \quad (2.58)$$

Similarly, for area means,

$$\begin{aligned} \underline{c}_{nn}^{\alpha} \alpha_{z,j} &= \frac{\sigma_n^2}{(2n+1)\Delta_{1j}} [\dots \int_{\theta_1}^{\theta_1+\Delta\theta} \bar{P}_{nn}(\cos \theta) \sin \theta d\theta \int_{\lambda_1}^{\lambda_1+\Delta\lambda} \underline{c}_n^{\alpha^T} d\lambda \dots]^T \\ &= \frac{\sigma_n^2}{(2n+1)\Delta_{1j}} [\dots \int_{\theta_1}^{\theta_1+\Delta\theta} \bar{P}_{nn}(\cos \theta) \sin \theta d\theta \left(\begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix} \underline{c}_n^0 + \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \underline{c}_n^1 \right)^T \dots]^T \quad (2.59) \end{aligned}$$

(where Δ_{1j} is supposed to be independent of j) according to expressions (1.7) and (2.7), so

$$\underline{f}_{nn}^{\alpha} = [\dots \chi_1^{nn} \left(\begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix} \underline{c}_n^0 + \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \underline{c}_n^1 \right)^T \dots]^T$$

In conclusion, the optimal estimator for point values has the form

$$\underline{c}_{nn}^{\alpha} = (\underline{f}_{nn}^{\alpha})^T \underline{m} = \sum_{i=0}^{Nr-1} \sum_{j=0}^{Nl-1} \chi_i^{nn} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta\lambda m_{1j} \quad (2.60)$$

while that for area means is of the type

$$\underline{c}_{nn}^{\alpha} = (\underline{f}_{nn}^{\alpha})^T \underline{m} = \sum_{i=0}^{Nr-1} \sum_{j=0}^{Nl-1} \chi_i^{nn} \begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix} \cos m_j \Delta\lambda + \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \sin m_j \Delta\lambda m_{1j} \quad (2.61)$$

so they are both of the quadratures kind, as anticipated in the preceding paragraph.

2.11 Setting up and Inverting the Covariance Matrix

Each block C^{1p} of $(C_{zz} + D)$ is wholly determined by the $1st \frac{1}{2}Nl + 1$ elements in its first row; if the number of operations required to compute any

element of C^{1p} is k , then only $(\frac{1}{2}Nl + 1)k$ operations are needed per block, instead of Nl^2k , as would be the case if C^{1p} did not have the Toeplitz structure described previously. This is a reduction of the number of operations by a factor of $2Nl$, and clearly applies not only to C^{1p} but to the whole covariance matrix as well. So $(C_{zz} + D)$ can be set up about $2Nl$ times faster than an ordinary matrix of the same size.

The total number of elements to be computed is $\frac{1}{2}Nl \times Nr^2$, or N^3 in the case of equal angular grids. If the grid is a fine one, this can still be a very large number of covariances. This is particularly serious in the case of area means, because the area mean covariances are given by expressions of the form

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{u}_{pq}) &= M \left\{ \int_{\sigma_{1j}} u d\sigma \int_{\sigma_{pq}} u d\sigma \right\} = \int_{\sigma_{1j}} \int_{\sigma_{pq}} M \left\{ u(\theta, \lambda) u(\theta', \lambda') \right\} d\sigma d\sigma' \\ &= \int_{\sigma_{1j}} \int_{\sigma_{pq}} \text{cov}(u(\theta, \lambda), u(\theta', \lambda')) d\sigma d\sigma' \end{aligned} \quad (2.62)$$

involving double area integrals of the covariance function. Numerical quadratures methods, such as the one described in paragraph (4.3), have been used in the past to obtain $\text{cov}(\bar{u}_{1j}, \bar{u}_{pq})$ (see, for example, Rapp (1977)). These methods take so much time in the case of fine equal angular grids for instance, that it may be practically impossible to use them to set up the covariance matrix of a global data set, in spite of the reduction by $2Nl$ in the number of operations. Fortunately, the coefficients a_n^{1p} in the Fourier expansion of the elements

$$c_{jq}^{1p} = \text{cov}(\bar{u}_{1j}, \bar{u}_{pq}) + E \{ n_{1j} n_{pq} \}$$

(expression (2.53)) can be obtained by means of a series expansion (truncated to a conveniently high degree N_{max}) according to expression (4.14) in paragraph (4.1). These coefficients are

$$a_n^{1p} = \left[\sum_{h=0}^K \sum_{n=n}^{N_{max}} I_{n, 2Nh+n, 1} I_{n, 2Nh+n, p} + \sum_{h=1}^K \sum_{n=n}^{N_{max}} I_{n, 2Nh-n, 1} I_{n, 2Nh-n, p} \right] F(m)$$

where

$$I_{n, z, 1} = \int_{\sigma_1}^{\sigma_1 + \Delta\theta} \bar{P}_{nz}(\cos\theta) \sin\theta d\theta \frac{\sigma_n}{\sqrt{2n+1}} \frac{1}{\Delta\lambda(\cos\theta_1 - \cos(\theta_1 + \Delta\theta))} \quad (2.63)$$

$$F(m) = \begin{cases} \Delta\lambda^2 & \text{if } m = 0 \\ (2m^{-2}) (1 - \cos m\Delta\lambda) & \text{and } (2h+1)N \leq N_{max} \end{cases}$$

A similar reasoning to that for area means leads to an analogous formula for point values:

$$\begin{aligned} a_n^{1p} &= \left[\sum_{h=0}^K \sum_{n=n}^{N_{max}} (\bar{P}_{n, 2Nh+n}(\cos\theta_1) \bar{P}_{n, 2Nh+n}(\cos\theta_p)) + \sum_{h=0}^K \sum_{n=n}^{N_{max}} (\bar{P}_{n, 2Nh-n}(\cos\theta_1) \right. \\ &\quad \left. \cdot \bar{P}_{n, 2Nh-n}(\cos\theta_p)) \right] \frac{\sigma_n^2}{2n+1} \end{aligned} \quad (2.64)$$

The importance of (2.63) and (2.64) is that, if the signal and noise are such that the number of terms in the summations is not too large (N_{max} is a "manageable" number), they allow the direct determination of the elements of the $R(m)$ matrices according to (2.54). In this way, the $R(m)$ can be created without first having to set up the whole covariance matrix and then to obtain the discrete Fourier transform of the first row of each $C^{(m)}$. This advantage further increases in the case when the grid is symmetrical with respect to the equator, a situation that applies to all equal angular grids. Then each $R(m)$ is persymmetrical, i.e., symmetrical with respect both to the main diagonal and the main antidiagonal, provided D is also persymmetrical (for instance, uniform noise). This means that only approximately $\frac{1}{2}N^2$ elements in each $R(m)$ are different and have to be calculated individually.

Having set up the $R(m)$ without first creating the covariance matrix, the inverse of $(C_{zz} + D)$ can be found by the equivalent operation of obtaining all $R(m)^{-1}$. The number of operations in a matrix inversion is usually $O(\text{dimension}^3)$, or $O(N^3)$ for a covariance matrix of an equal angular data set. The number of operations per $R(m)$ is $O(N^3)$, or $O(N^3)$ for the equal angular grid. In fact, as explained in (Colombo, 1979a), the inversion of a persymmetrical $R(m)$ is equivalent to that of two matrices of half its dimension, one related to vectors of frequency m of the cosine type, and the other to vectors of the same frequency of the sine type. This further reduces calculation by a factor of $\frac{1}{2}$. With $O(N)$ $R(m)$ matrices to be inverted, the total comes to $O(N^4)$ operations, or $O(N^2)$ times less than for the inversion of $(C_{zz} + D)$ by ordinary techniques (Choleskii factorization, Gauss-Jordan elimination, etc.). $O(N^2)$ is also the order of the number of data points in the grid, so in the case of a $1^\circ \times 1^\circ$ equal angular grid with 64800 elements the reduction in computing time is $O(64800)$.

The numerical examples in section 3 all involve $5^\circ \times 5^\circ$ data sets with 2592 elements, so $(C_{zz} + D)$ is of dimension 2592. Setting up and inverting such a matrix is a large exercise, even with a modern digital computer such as the AMDHAL 470 at Ohio State, unless the matrix has a strong structure that can be exploited to simplify the work. As such is indeed the case here, the subroutine NORMAL described in Appendix B has been able to do the whole setting up and inversion in only 20 seconds.

The inversion of $(C_{zz} + D)$ requires $O(\text{dimension}^3)$ operations ($O(N^4)$) instead of $O(\text{dimension}^3)$ because of the Toeplitz circulant structure of the $C^{(m)}$ blocks. This " $O(\text{dimension}^2)$ " property is common to other algorithms for inverting Toeplitz-type matrices, such as the famous Trench algorithm (Trench, 1965), and the Justice algorithm (Justice, 1977), the first for data sampled on the real line and the second for data sampled on the plane. So, in spite of its "rather wicked" nature, the sphere allows this very convenient property of regular grids to apply also on its surface. In fact, not only on the sphere, but also on any body of revolution (cone, oblate and prolate spheroid, hyperboloids and paraboloids of revolution, etc.) regular sampling and

covariances that satisfy the " $|j-q|$ condition" will result in covariance matrices of the type described here, and this is also true of other matrices based on symmetrical kernels, such as the normal matrices of point mass models, when the points belong to a regular grid, etc. Finally, the optimal estimator $\hat{C}_{nn}^\alpha = \hat{f}_{nn}^\alpha \underline{m}$ for this type of covariance matrix is, as shown in the previous paragraph, of the quadratures type, so the optimal \hat{C}_{nn}^α can be obtained using the same efficient algorithms described in paragraphs (1.5)-(1.7). Altogether, the powerful structure of the covariance matrix for regular global data sets is most remarkable. One of its many advantageous features is that, because the creation and inversion of each $R(m)$ can be done quite independently from those of the others, the algorithms developed for this type of matrices are eminently suited for implementation in parallel processing computers.

The separation of the algorithm according to orders also means that, although setting up and inverting all the $R(m)$ may require a large number of operations, only a fraction of those actually correspond to the recovery of the \hat{C}_{nn}^α of any given m , so the numerical errors due to rounding or truncation are not likely to accumulate to any great extent in the results.

2.12 Optimal Formulas for non-Uniform, Correlated Noise

Irregular noise, already discussed in paragraph (2.9), may be due not only to the varying quality of the measurements, but also to the way the data is "grided", i.e., the way the value attributed to a node (or block) ij is obtained by interpolation from actual measurements nearby, as usually data is not sampled regularly on a global basis. As the number, disposition, and quality of the measurements used will vary from point to point in the grid, so will the accuracy of the interpolated values. Furthermore, even if the measurements themselves are not correlated, the grided values may be correlated because some of the data may be used for more than one interpolated value. This brings about the question of what can be done when D is neither diagonal, nor are the D^{ir} blocks in D , corresponding to the C^{ir} blocks in C_{zz} , all Toeplitz circulant. The answer is a simple extension of the results already obtained for the uncorrelated case.

When the noise is both non-stationary and correlated, replacing the covariances $E \{n_{1j} n_{rs}\}$ with

$$\bar{\sigma}^{ir|h|} = \frac{1}{4N} \left[\sum_{j=0}^{2N-1} (E \{n_{1j} n_{r, j+h}\} + E \{n_{1j} n_{r, j-h}\}) \right], \text{ where } h = j - s,$$

will result in a modified "noise matrix" L where the L^{ir} (corresponding to the D^{ir} and the C^{ir}) will be all Toeplitz circulant, because the "covariance" $\bar{\sigma}^{ir|h|}$ satisfies the condition that, for a given i and r , it is a function of $|j-s|$ alone. The optimal estimator for the modified measure

$$\hat{\Phi}(L) = \text{tr}[C - 2C_{zz} F^T + F(C_{zz} + L) F]$$

The importance of (2.63) and (2.64) is that, if the signal and noise are such that the number of terms in the summations is not too large ($N_{\alpha\alpha}$ is a "manageable" number), they allow the direct determination of the elements of the $R(m)$ matrices according to (2.54). In this way, the $R(m)$ can be created without first having to set up the whole covariance matrix and then to obtain the discrete Fourier transform of the first row of each C^b . This advantage further increases in the case when the grid is symmetrical with respect to the equator, a situation that applies to all equal angular grids. Then each $R(m)$ is persymmetrical, i.e., symmetrical with respect both to the main diagonal and the main antidiagonal, provided D is also persymmetrical (for instance, uniform noise). This means that only approximately $\frac{1}{4}N^2$ elements in each $R(m)$ are different and have to be calculated individually.

Having set up the $R(m)$ without first creating the covariance matrix, the inverse of $(C_{zz} + D)$ can be found by the equivalent operation of obtaining all $R(m)^{-1}$. The number of operations in a matrix inversion is usually $O(\text{dimension}^3)$, or $O(N^6)$ for a covariance matrix of an equal angular data set. The number of operations per $R(m)$ is $O(N^3)$, or $O(N^3)$ for the equal angular grid. In fact, as explained in (Colombo, 1979a), the inversion of a persymmetrical $R(m)$ is equivalent to that of two matrices of half its dimension, one related to vectors of frequency m of the cosine type, and the other to vectors of the same frequency of the sine type. This further reduces calculation by a factor of $\frac{1}{4}$. With $O(N)$ $R(m)$ matrices to be inverted, the total comes to $O(N^4)$ operations, or $O(N^2)$ times less than for the inversion of $(C_{zz} + D)$ by ordinary techniques (Choleskii factorization, Gauss-Jordan elimination, etc.). $O(N^2)$ is also the order of the number of data points in the grid, so in the case of a $1^\circ \times 1^\circ$ equal angular grid with 64800 elements the reduction in computing time is $O(64800)$.

The numerical examples in section 3 all involve $5^\circ \times 5^\circ$ data sets with 2592 elements, so $(C_{zz} + D)$ is of dimension 2592. Setting up and inverting such a matrix is a large exercise, even with a modern digital computer such as the AMDHAL 470 at Ohio State, unless the matrix has a strong structure that can be exploited to simplify the work. As such is indeed the case here, the subroutine NORMAL described in Appendix B has been able to do the whole setting up and inversion in only 20 seconds.

The inversion of $(C_{zz} + D)$ requires $O(\text{dimension}^3)$ operations ($O(N^4)$) instead of $O(\text{dimension}^3)$ because of the Toeplitz circulant structure of the C^b blocks. This " $O(\text{dimension}^2)$ " property is common to other algorithms for inverting Toeplitz-type matrices, such as the famous Trench algorithm (Trench, 1965), and the Justice algorithm (Justice, 1977), the first for data sampled on the real line and the second for data sampled on the plane. So, in spite of its "rather wicked" nature, the sphere allows this very convenient property of regular grids to apply also on its surface. In fact, not only on the sphere, but also on any body of revolution (cone, oblate and prolate spheroid, hyperboloids and paraboloids of revolution, etc.,) regular sampling and

covariances that satisfy the " $|j-q|$ condition" will result in covariance matrices of the type described here, and this is also true of other matrices based on symmetrical kernels, such as the normal matrices of point mass models, when the points belong to a regular grid, etc. Finally, the optimal estimator $\hat{C}_{nn}^\alpha = \hat{f}_{nn}^{\alpha} \underline{m}$ for this type of covariance matrix is, as shown in the previous paragraph, of the quadratures type, so the optimal \hat{C}_{nn}^α can be obtained using the same efficient algorithms described in paragraphs (1.5)-(1.7). Altogether, the powerful structure of the covariance matrix for regular global data sets is most remarkable. One of its many advantageous features is that, because the creation and inversion of each $R(m)$ can be done quite independently from those of the others, the algorithms developed for this type of matrices are eminently suited for implementation in parallel processing computers.

The separation of the algorithm according to orders also means that, although setting up and inverting all the $R(m)$ may require a large number of operations, only a fraction of those actually correspond to the recovery of the \hat{C}_{nn}^α of any given m , so the numerical errors due to rounding or truncation are not likely to accumulate to any great extent in the results.

2.12 Optimal Formulas for non-Uniform, Correlated Noise

Irregular noise, already discussed in paragraph (2.9), may be due not only to the varying quality of the measurements, but also to the way the data is "grided", i.e., the way the value attributed to a node (or block) ij is obtained by interpolation from actual measurements nearby, as usually data is not sampled regularly on a global basis. As the number, disposition, and quality of the measurements used will vary from point to point in the grid, so will the accuracy of the interpolated values. Furthermore, even if the measurements themselves are not correlated, the grided values may be correlated because some of the data may be used for more than one interpolated value. This brings about the question of what can be done when D is neither diagonal, nor are the D^{ir} blocks in D , corresponding to the C^{ir} blocks in C_{zz} , all Toeplitz circulant. The answer is a simple extension of the results already obtained for the uncorrelated case.

When the noise is both non-stationary and correlated, replacing the covariances $E\{n_{1j}, n_{r,s}\}$ with

$$\bar{\sigma}^{ir|h|} = \frac{1}{4N} \left[\sum_{j=0}^{2N-1} (E\{n_{1j}, n_{r,j+h}\} + E\{n_{1j}, n_{r,j-h}\}) \right], \text{ where } h = j - s,$$

will result in a modified "noise matrix" L where the L^{ir} (corresponding to the D^{ir} and the C^{ir}) will be all Toeplitz circulant, because the "covariance" $\bar{\sigma}^{ir|h|}$ satisfies the condition that, for a given i and r , it is a function of $|j-s|$ alone. The optimal estimator for the modified measure

$$\hat{\Phi}(L) = \text{tr}[C - 2C_{zz} F^T + F(C_{zz} + L) F]$$

must be of the quadratures type, because of the structure of L . To show that is also the best estimator of this kind in terms of the original norm

$$\hat{\phi}(D) = \text{tr} [C - 2C_{cz} F^T + F(C_{zz} + D) F^T],$$

the proof will proceed much as in the case of paragraph (2.9).

The propagated error measure for \hat{C}_{zz}^α is

$$\begin{aligned} \sigma_{\eta, n}^{2\alpha} &= E \left\{ \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \chi_i^{nn} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda n_{1j} \sum_{r=0}^{N-1} \sum_{s=0}^{2N-1} \chi_r^{nn} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_s \Delta \lambda n_{rs} \right\} \\ &= \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \sum_{r=0}^{N-1} \sum_{s=0}^{2N-1} \chi_i^{nn} \chi_r^{nn} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_s \Delta \lambda E \{n_{1j} n_{rs}\} \end{aligned}$$

so, for \hat{C}_{zz} and \hat{S}_{zz} combined,

$$\sum_{\alpha=0}^1 \sigma_{\eta, n}^{2\alpha} = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \sum_{r=0}^{N-1} \sum_{s=0}^{2N-1} \chi_i^{nn} \chi_r^{nn} \cos m(j-s) \Delta \lambda E \{n_{1j} n_{rs}\}$$

thus

$$\begin{aligned} \hat{\phi}_{\eta}(D) &= \text{tr} [F D F^T] = \sum_{n=0}^{N-1} \sum_{j=0}^{2N-1} \sum_{i=0}^{N-1} \sum_{r=0}^{N-1} \chi_i^{nn} \chi_r^{nn} \sum_{j=0}^{2N-1} \sum_{s=0}^{2N-1} \cos m(j-s) \Delta \lambda E \{n_{1j} n_{rs}\} \\ &= \sum_{n=0}^{N-1} \sum_{h=0}^{2N-1} \sum_{i=0}^{N-1} \sum_{r=0}^{N-1} \chi_i^{nn} \chi_r^{nn} \sum_{h=0}^{2N-1} \cos m h \Delta \lambda \sum_{j=0}^{2N-1} (E \{n_{1j} n_{r, j+h}\} + E \{n_{1j} n_{r, j-h}\}) \end{aligned}$$

Replacing both $E \{n_{1j} n_{r, j+h}\}$ and $E \{n_{1j} n_{r, j-h}\}$ in the last expression with $\bar{\sigma}^{1r|h|}$ is the same as replacing D with L , so

$$\begin{aligned} \hat{\phi}_{\eta}(L) &= \text{tr} [F L F^T] = \sum_{n=0}^{N-1} \sum_{j=0}^{2N-1} \sum_{i=0}^{N-1} \sum_{r=0}^{N-1} \chi_i^{nn} \chi_r^{nn} \sum_{h=0}^{2N-1} \cos m h \Delta \lambda \sum_{j=0}^{2N-1} 2 \bar{\sigma}^{1r|h|} \\ &= \sum_{n=0}^{N-1} \sum_{h=0}^{2N-1} \sum_{i=0}^{N-1} \sum_{r=0}^{N-1} \chi_i^{nn} \chi_r^{nn} \sum_{h=0}^{2N-1} \cos m h \Delta \lambda 4N \bar{\sigma}^{1r|h|} \\ &= \sum_{n=0}^{N-1} \sum_{h=0}^{2N-1} \sum_{i=0}^{N-1} \sum_{r=0}^{N-1} \chi_i^{nn} \chi_r^{nn} \sum_{h=0}^{2N-1} \cos m h \Delta \lambda \sum_{j=0}^{2N-1} (E \{n_{1j} n_{r, j+h}\} + E \{n_{1j} n_{r, j-h}\}) \end{aligned}$$

because of the definition of $\bar{\sigma}^{1r|h|}$. Comparing the expressions for $\hat{\phi}_{\eta}(L)$ and for $\hat{\phi}_{\eta}(D)$ it is clear that they are identical. From this follows that the modified error measure $\hat{\phi}(L)$ coincides with $\hat{\phi}(D)$ when the estimator is the optimal estimator of the quadratures type for $\hat{\phi}(L)$, and that this must be the optimal estimator of the quadratures type for $\hat{\phi}(D)$ as well. The other conclusions arrived at in paragraph (2.9) for the uncorrelated case apply equally well here.

2.13 Least Squares Adjustment, and Least Squares Collocation

(a) Band-Limited Signal

If there is a degree N_{max} above which the degree variances σ_n^2 are all negligible or zero, then the signal can be said to be band limited, and the data will satisfy equations of the type

$$m_{ij} = \sum_{n=0}^{N_{\max}} \sum_{s=0}^n \sum_{\alpha=0}^1 \bar{C}_{nn}^{\alpha} \bar{Y}_{nn}^{\alpha}(\theta_i, \lambda_j) + n_{ij} \quad (2.65)$$

(the treatment here is for point values; the extension to area means is trivial)
With one equation such as (2.65) per point in the grid, the result is a system of equations

$$\underline{m} + \underline{v} = A \underline{c} \quad (\underline{v} = -\underline{n}) \quad (2.66)$$

where A is a $N_p \times N_c$ matrix (N_p is the number of points in the grid, N_c the number of coefficients). The columns of A consist of elements of the type

$$a_{ij}^{nn} = \bar{Y}_{nn}(\theta_i, \lambda_j) \quad (2.67)$$

According to the discussion in paragraph (1.3), if the grid is equal angular, A has full rank when $N_{\max} < N$. $N_c = N^2$ and the upper limit in the summations is $N-1$ in what follows.

Least squares adjustment is a method for solving for the \bar{C}_{nn}^{α} while minimizing the propagated noise defined in paragraph (2.4). The least squares solution is

$$\begin{aligned} \underline{\hat{c}} &= (A^T D^{-1} A)^{-1} A^T D^{-1} \underline{m} \\ &= G^{-1} A^T D^{-1} \underline{m} \end{aligned} \quad (2.68)$$

where

$$G = A^T D^{-1} A \quad (2.69)$$

is the $N_c \times N_c$ normal matrix, while $D = E\{\underline{n} \underline{n}^T\}$ is the same noise matrix considered before. Clearly, the least squares estimator matrix is

$$F_{ls} = (A^T D^{-1} A) A^T D^{-1}$$

When the noise has zero mean ($E\{\underline{n}\} = \underline{0}$), the estimator of (2.68) is the best linear unbiased estimator, because it minimizes

$$\text{tr} [E\{F \underline{n} \underline{n}^T F^T\}] = \text{tr} [F D F^T]$$

and $E\{F(\underline{z} + \underline{n})\} = \underline{c}$. If, in addition to all this, the probability distribution of the noise is Gaussian, then (2.68) corresponds to the maximum likelihood estimator as well. In many scientific applications the noise has approximately zero mean and near-Gaussian distribution, while D is known reasonably well;

for this reason, methods based on expression (2.68) are used quite often. The linearity of the resulting estimators is helpful, because this avoids the use of methods based on non-linear formulas that are usually difficult both from a theoretical and from a practical point of view. The variances of the estimates are given by the corresponding diagonal elements of the a posteriori variance-covariance matrix

$$E_{\underline{c}_s} = (A^T D^{-1} A)^{-1} = G^{-1}$$

Therefore, to obtain both the estimates and their variances it is necessary to know G^{-1} . Sometimes, because of the nature of A and D , G can be seriously ill-conditioned, the inversion suffering from strong numerical instabilities. To reduce this problem, a simple device known as regularization is often used (see, for instance, Tikhonov and Arsenin, 1977). Generally speaking, regularization is the introduction of a slight change in a problem, so the solution virtually remains the same, but the modified problem has better numerical properties. In least squares methods regularization usually implies adding a small positive definite matrix K (diagonal, as a rule) to G before attempting to invert it. The regularized optimal estimator would be

$$\underline{\hat{c}} = (A^T D^{-1} A + K)^{-1} A^T D^{-1} \underline{m} \quad (2.70)$$

The inverse of the covariance matrix of the harmonic coefficients C is a positive, diagonal matrix which could be used to regularize the normal matrix:

$$\underline{\hat{c}} = (A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1} \underline{m} \quad (2.71)$$

It is easy to see that this expression minimizes the quadratic form

$$Q = \underline{c}^T C^{-1} \underline{c} + \underline{v}^T D^{-1} \underline{v} \quad (2.72)$$

subject to the constraint

$$\underline{m} = A \underline{c} + \underline{v} \quad (2.73)$$

Moreover, (2.72) is the equivalent of the least squares collocation error measure when the signal is band-limited (Moritz, (1980)). This idea has been used, among others, by Schwarz (1975) for the determination of low degree zonal coefficients of the geopotential, and by Lerch et al. (1979), who employed it to stabilize the adjustment of the GEM-9 gravity field model with remarkable success. The equivalence of (2.71) to the collocation estimator is true only for band-limited signals; in the "real world" the gravity field has infinite bandwidth, so (2.71) is no more than an approximation. The band-limited assumption is a reasonable one, however, as the σ_n eventually become negligible for large n . This is particularly true at satellite altitudes; in any case, geodesy is a science of wise approximations. Moritz (1980) has provided a very clear and concise explanation of the use of collocation in general, and expression (2.71) in particular, in spherical harmonic analysis.

An alternative derivation of (2.71) follows from the matrix equation

$$C A^T (A C A^T + D)^{-1} = (A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1} \quad (2.74)$$

(see, for instance, Uotila (1978), equation (29)), which is valid for symmetrical matrices, provided the inverses or pseudoinverses of D , C , and $(A C A^T + D)$ do exist. According to the definitions in paragraph (2.2):

$$C_{zz} = M\{\underline{c} \underline{z}^T\} = M\{\underline{c} \underline{c}^T A^T\} = M\{\underline{c} \underline{c}^T\} A^T = C A^T \quad (2.75)$$

$$C_{zz} = M\{\underline{z} \underline{z}^T\} = M\{A \underline{c} \underline{c}^T A^T\} = A M\{\underline{c} \underline{c}^T\} A^T = A C A^T \quad (2.76)$$

Replacing C_{zz} and C_{zz} in the expression of the collocation estimator matrix (2.39) with their equivalents given by (2.75) and (2.76):

$$F = C_{zz} (C_{zz} + D)^{-1} = C A^T (A C A^T + D)^{-1} = (A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1} \quad (2.77)$$

according to equation (2.74). This shows that the "regularized" estimator matrix $(A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1}$ is indeed the same as the collocation estimator matrix, so (2.71) represents an alternative form of collocation when the data is band-limited.

(b) Infinite Bandwidth

In this case the "observation equations" are

$$m_{1j} = \sum_{n=0}^{\infty} \sum_{s=0}^n \sum_{\alpha=0}^1 \bar{C}_{ns}^{\alpha} \bar{Y}_{ns}^{\alpha}(\theta_1, \lambda_1) + n_{1j}$$

Calling

$$w_{1j} = \sum_{n=N+1}^{\infty} \sum_{s=0}^n \sum_{\alpha=0}^1 \bar{C}_{ns}^{\alpha} \bar{Y}_{ns}^{\alpha}(\theta_1, \lambda_1) \quad (2.78a)$$

and

$$\underline{w} = [w_1 \dots w_k \dots w_{Np}]^T \quad (k = 2Ni + j) \quad (2.78b)$$

then

$$m_{1j} = \sum_{n=0}^N \sum_{s=0}^n \sum_{\alpha=0}^1 \bar{C}_{ns}^{\alpha} \bar{Y}_{ns}^{\alpha}(\theta_1, \lambda_1) + w_{1j} + n_{1j} \quad (2.79)$$

and, regarding this expression as a modified observation equation, and replacing D with $D + M\{\underline{w} \underline{w}^T\}$ in (2.71), the linear estimator that minimizes the quadratic form

$$\tilde{Q} = \underline{c}^T C^{-1} \underline{c} + \underline{v}^T (D + M\{\underline{w} \underline{w}^T\})^{-1} \underline{v} \quad (2.80)$$

is

$$\hat{\underline{c}} = (A^T (D + M\{\underline{w} \underline{w}^T\})^{-1} A + C^{-1})^{-1} A^T (D + M\{\underline{w} \underline{w}^T\})^{-1} \underline{m} \quad (2.81)$$

It is easy to show, either following the lines of Moritz (1972), or going

back once more to the matrix identity (2.74), that expression (2.81) is identical with the estimator of least squares collocation. Rigorously speaking, expression (2.81) should be used whenever $\sigma_n^2 \neq 0$ for $n > N$, even if $\sigma_n^2 = 0$ for $n > N_{max}$, for some finite $N_{max} > N$.

2.14 Ridge Regression and Least Squares Collocation

Consider once more the estimator

$$\hat{\underline{c}} = F_{\ell} \underline{m} = (A^T D^{-1} A)^{-1} A^T D^{-1} \underline{m}$$

If there is no noise, so $\underline{m} = \underline{z} = A \underline{c}$, and if $n < N$, then

$$\hat{\underline{c}} = (A^T D^{-1} A)^{-1} A^T D^{-1} A \underline{c} = \underline{c}$$

According to the definition given in paragraph (2.4), the sampling error of this estimator is zero, so the measure of this error must be also zero. If the noise has zero mean, it follows that

$$E\{\hat{\underline{c}}\} = E\{F_{\ell} \underline{m}\} = F_{\ell} E\{A \underline{c} + \underline{n}\} = F_{\ell} A \underline{c} + E\{\underline{n}\} = \underline{c}$$

or, as it is usually said, the estimator is unbiased. Moreover, by a simple extension of the Gauss-Markov theorem to the case of a general symmetrical positive matrix D (see, for instance, Bibby and Toutenberg, 1977), $F_{\ell} \underline{m}$, of all linear unbiased estimators, has the least propagated error measure $\text{tr}\{F_{\ell} D F_{\ell}^T\} = \text{tr}\{(A^T D^{-1} A)^{-1}\}$, as mentioned previously.

The estimator of expression (2.77) does not, in general, give perfect estimates of \underline{c} in the absence of noise: it is a biased estimator, and the measure of the bias is $\text{tr}\{C - 2C_{oz} F^T + F(C_{zz} + D) F^T\}$ (this term¹⁾ can no longer be regarded as the measure of the sampling error, as it is the presence of C^{-1} inside the parenthesis in (2.71) and not the sampling that brings about this error). According to (2.39), F is the estimator that minimizes the total error measure, so

$$\begin{aligned} \sigma_c^2 &= \text{tr}\{C - 2C_{oz} F^T + F(C_{zz} + D) F^T\} \\ &\leq \text{tr}\{C - 2C_{oz} F_{\ell}^T + F_{\ell}(C_{zz} + D) F_{\ell}^T\} \\ &= \text{tr}\{F_{\ell} D F_{\ell}^T\} \end{aligned}$$

If the covariance matrix is positive definite (which only requires that all $\sigma_n^2 \neq 0$ for $0 \leq n < N$) then the last expression applies with strict inequality

$$\begin{aligned} \text{tr}\{C - 2C_{oz} F^T + F(C_{zz} + D) F^T\} &< \text{tr}\{F_{\ell} D F_{\ell}^T\} \\ &= \text{tr}\{(A^T D^{-1} A)^{-1}\} \end{aligned}$$

¹ In the band-limited case.

Some may find this result rather surprising: the best estimator with zero bias is in fact worse than the biased estimator of (2.71)! The difficulty is only apparent: F_{β} is the best estimator with no bias; once the "no bias" condition is removed, the expression above merely indicates that there is an estimator in the larger class of the estimators that have a bias (including those with zero bias) such that the sum of the bias and the propagated noise is smaller. From this, it is clear that

$$\text{tr} \{F D F^T\} < \text{tr} \{(A^T D^{-1} A)\} \quad (\text{as } \text{tr} \{C - 2C_{cz} F^T + F C_{zz} F^T\} \geq 0)$$

which is indeed possible when the condition $C - 2C_{cz} F^T + F C_{zz} F^T = 0$ is removed. In fact, there is nothing very new about all this: the use of biased estimators to obtain estimates with small variances is a reasonably well-established practice in applied statistics. In particular, the technique known as ridge regression consists in using the biased estimator

$$\underline{\hat{c}} = (\chi^T \chi + K)^{-1} \chi^T \underline{m}$$

with a suitable choice of K (Bibby, 1972). Clearly, this expression is the same as (2.71) when $\chi = A$, $D=I$, and $K=C^{-1}$. With some obvious modifications suggested by (2.81), this argument can be extended to the estimation of \underline{c} when $n \geq N$, so it can be said that within the scope of spherical harmonic analysis least squares collocation is a form of ridge regression.

This brings up the question of just how realistic the error measure is; after all the best of all possible estimators in terms of a given norm could be a very bad one for some specific problem where that norm is not suitable. To answer this question, one must start by defining the meaning of "realistic". If one is interested in minimizing the actual error variance of the coefficients per degree, i.e., the expression

$$\delta_n^2 = \sum_{\alpha=0}^l \sum_{n=0}^n (\hat{C}_{n\alpha}^\alpha - \bar{C}_{n\alpha}^\alpha)^2 (2n+1)^{-1}$$

which may be of interest because this corresponds to, say, the global mean square of the error of representing the continuous function $(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{\alpha=0}^n \bar{C}_{n\alpha}^\alpha Y_{n\alpha}(\theta, \lambda)$ with $f(\theta, \lambda) = \sum_{n=0}^{\infty} \sum_{\alpha=0}^n \hat{C}_{n\alpha}^\alpha Y_{n\alpha}(\theta, \lambda)$, according to Parseval's Identity (exp. (1.12)), then one could say that a realistic measure is one that gives close estimates of the actual error variances. The "actual error" measure δ_n^2 corresponds to one of infinitely many "events" over which the collocation measure is an average. The proof of a pudding being in the eating, the reader can judge just how realistic the collocation measure is by looking at the numerical results in section 3, where the

value of the measure turns out to differ only by a small percentage from the actual variance in each one of a number of simulated "events", i.e., the recovery of the \bar{C}_{nn}^α from "simulated data", where the \bar{C}_{nn}^α are known random numbers scaled to have the desired power spectrum.

2.15 Structure of the Normal Matrix

The elements of the normal matrix G (in the case of point data) are of the form

$$\begin{aligned} g_{nn,pq}^{\alpha,\beta} &= \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{Y}_{nn}^\alpha(\theta_i, \lambda_j) \bar{Y}_{pq}^\beta(\theta_i, \lambda_j) \sigma_i^{-2} \\ &= \sum_{i=0}^{N-1} \bar{P}_{nn}(\cos \theta_i) \bar{P}_{pq}(\cos \theta_i) \sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} q_j \Delta \lambda \sigma_i^{-2} \end{aligned} \quad (2.82)$$

where $\sigma_i^2 = \sigma_1^2$, for all $0 \leq j \leq 2N$ (i.e., "regular" noise as defined in paragraph 2.8). If $n, m < N$, then the following equations apply:

$$\sum_{j=0}^{2N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} q_j \Delta \lambda = 0 \quad \text{if} \quad \begin{cases} \alpha \neq \beta \\ \text{or} \\ m \neq q \end{cases} \quad (2.83)$$

Moreover, if the grid is symmetrical with respect to the Equator (as equal angular grids are), and if $\sigma_1^2 = \sigma_{N-1}^2$ (for instance, if σ_1^2 is independent also of i) the relationships

$$\sigma_1^{-2} \bar{P}_{nn}(\cos \theta_i) = \bar{P}_{nn}(\cos \theta'_i) (-1)^{n-1} \sigma_{N-1}^{-2}$$

$$\sigma_1^{-2} \bar{P}_{pn}(\cos \theta_i) = \bar{P}_{pn}(\cos \theta'_i) (-1)^{p-1} \sigma_{N-1}^{-2}$$

$$\theta'_i = \pi - \theta_i$$

must apply, according to par. (1.2), and from these follows

$$\sum_{i=0}^{N-1} \sigma_i^{-2} \bar{P}_{nn}(\cos \theta_i) \bar{P}_{pn}(\cos \theta_i) = 0 \quad \text{if } n - p \text{ is odd.} \quad (2.84)$$

In brief: if $n, m < N$, and the grid is regularly spaced in longitude, and symmetrical with respect to the equator, then

$$g_{nn,pq}^{\alpha,\beta} = 0 \quad \text{if} \quad \begin{cases} \alpha \neq \beta, m \neq q \\ \text{or} \\ n - p \text{ is odd} \end{cases} \quad (2.85)$$

If neither of the conditions listed above apply, then $g_{nn,pq}^{\alpha,\beta}$ may or may not be zero. If the coefficients \bar{C}_{nn}^α are ordered in \underline{c} so that all those of the same m are grouped together, and for a given m all \bar{C}_{nn}^1 are separated from all \bar{C}_{nn}^0 , and further more all \bar{C}_{nn}^α with $n - m$ even are separated from all those with $n - m$ odd, then the normal matrix becomes arranged in such a way that all potentially non-zero elements (i.e., not satisfying (2.85)) are also grouped together forming a series of diagonal blocks $G_n^{\alpha,\delta}$, where δ signals

the parity of $n-p$. Each one of these diagonal blocks is made exclusively of one of the following types of elements

$$\begin{cases} g_{n,p}^{0,0} & \left\{ \begin{array}{l} n-p \text{ even } (\delta=0) \\ n-p \text{ odd } (\delta=1) \end{array} \right. \\ g_{n,p}^{1,1} \end{cases}$$

If the grid is not symmetrical with respect to the equator, groups of type $\begin{pmatrix} \alpha=0 \\ \delta=0 \end{pmatrix}$ and $\begin{pmatrix} \alpha=0 \\ \delta=1 \end{pmatrix}$, or $\begin{pmatrix} \alpha=1 \\ \delta=0 \end{pmatrix}$ and $\begin{pmatrix} \alpha=1 \\ \delta=1 \end{pmatrix}$, become included into larger non-zero blocks, as there are more non-zero elements in that case. However, here the discussion will cover only the symmetrical case.

The largest blocks are $G_0^{\alpha,\delta}$, and their dimension is N_p^2 ; the smallest blocks have dimension 1×1 , for example $G_{N-1}^{\alpha,\delta}$.

The inverse of any block-diagonal matrix such as G is another block-diagonal matrix made up of the inverse of the blocks of G . There are $4N-2$ diagonal blocks $G_n^{\alpha,\delta}$ in G , and as many in G^{-1} .

The eigenvalues of G are those of the diagonal blocks, and the eigenvectors of G , those of the same blocks "expanded" with zeroes at both ends, so as to reach the dimension N_p of G .

The estimates' vector

$$\underline{\hat{c}} = G^{-1} A^T D^{-1} \underline{m}$$

can be partitioned in the same way as \underline{c} , each partition $\underline{\hat{c}}_n^{\alpha,\delta}$ including all coefficients' estimates of the same m , α , and parity δ of $n-m$:

$$\underline{\hat{c}}_n^{\alpha,\delta} = (G_n^{\alpha,\delta})^{-1} A_n^{\alpha,\delta} D^{-1} \underline{m} \quad (2.86)$$

where $A_n^{\alpha,\delta}$ is a $(N-m) \times N_p$ matrix with rows that are N_p -vectors of the type

$$\underline{a}_{n,n}^{\alpha,\delta} = [\dots \bar{P}_{nm}(\cos \theta_1) \left\{ \begin{array}{c} \cos \\ \sin \end{array} \right\} m_j \Delta \lambda \dots] \quad (2.87)$$

($n-m$ even if $\delta=0$, odd if $\delta=1$) So the rows of $(G_n^{\alpha,\delta})^{-1} A_n^{\alpha,\delta}$ are linear combinations of vectors of the same frequency m and, therefore, also vectors of the same frequency:

$$\underline{h}_{n,n}^{\alpha} = [\dots \chi_{1n}^{\alpha} \left\{ \begin{array}{c} \cos \\ \sin \end{array} \right\} m_j \Delta \lambda \dots] \quad (2.88)$$

Consequently, the estimate of a given \bar{c}_{nn}^{α} is

$$\hat{c}_{n,n}^{\alpha} = \underline{h}_{n,n}^{\alpha} D^{-1} \underline{m} = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \chi_{1n}^{\alpha} \left\{ \begin{array}{c} \cos \\ \sin \end{array} \right\} m_j \Delta \lambda \sigma_i^{-2} m_j \quad (2.89)$$

and this is a quadratures type estimator.

Setting-up and inverting the $G_{\alpha\delta}$ blocks is tantamount to setting up and inverting the whole matrix G . Since all operations related to one of the blocks are independent from those for the others, the inversion of G is ideally suited for parallel-processing computing. On average, each $G_{\alpha\delta}$ requires $O(N^3)$ operations to invert, or $O(N^4)$ altogether. Inverting G by ordinary techniques would involve $O((N^2)^3) = O(N^6)$, so there is an increase in efficiency of $O(N^2)$.

Finally, it is quite simple to show that these properties carry over both to the case of area means, and to problems where the surface being studied is not a sphere, but a surface of revolution symmetrical about a plane perpendicular to the axis of rotation, provided that the longitude increments be constant and the grid symmetrical with respect to the "equator". In this latter case, the expansion of the signal in solid spherical harmonics is

$$z_{ij} = \sum_{l=0}^{N-1} \sum_{m=0}^{2N-1} \sum_{\alpha=0}^l \frac{a_{lm}^{\alpha}}{r_{ij}^{l+1}} \bar{Y}_{lm}^{\alpha}(\theta_i, \lambda_j)$$

and the factors $\frac{a_{lm}^{\alpha}}{r_{ij}^{l+1}}$ are symmetrical about the equator, from which all the properties already mentioned for G follow.

Clearly, the structure of G possesses many properties similar or identical to those of $(C_{zz} + D)$ when the data is regularly sampled on a surface of revolution. These similitudes underline the intimate relationship between least squares adjustment and least squares collocation shown in the preceding paragraph. In fact, as least squares, regularized least squares, and collocation differ only in the diagonal matrix $(K, \text{ or } C^{-1})$ being added to G in expressions (2.68), (2.70), and (2.77), all the properties mentioned here for G apply to the normal matrices in each of the three methods equally well. The one important consideration, in the case of collocation, is that the data be band-limited. Otherwise, expression (2.81) indicates that $(C + M\{\underline{w} \underline{w}^T\})^{-1}$ and not C^{-1} must be added to G . Matrix $C + M\{\underline{w} \underline{w}^T\}$ has the same Toeplitz-type structure of $(C_{zz} + D)$ discussed in paragraph (2.10). Therefore, creating and inverting the normal matrix requires: (a) creating and inverting $C + M\{\underline{w} \underline{w}^T\}$, and (b) creating and inverting $A^T D^{-1} A + (M\{\underline{w} \underline{w}^T\} + C)^{-1}$, which can be shown to have the same block diagonal structure discussed here. This is twice the work needed to set up and invert $(C_{zz} + D)$ using the approach of paragraph (2.11), so, in the case of infinite bandwidth, that approach is more economical in computing and, therefore, more practical.

There may be one important point in favor of using formula (2.77) or (2.81) rather than formula (2.39) for obtaining the optimal estimator matrix F at least in the band-limited case: as the density of the grid increases, matrix $(C_{zz} + D)$ becomes increasingly more ill-conditioned, because the closer distance between data points results in covariances that have much the same values in consecutive rows or columns. On the other hand, the non-zero diagonal blocks in G are likely to become more and more diagonal-dominant as $\Delta\theta, \Delta\lambda \rightarrow 0$. This will depend on D : for instance, if the variances of the noise were of the form

$$\sigma_{11}^2 = \sin \theta_1$$

then

$$\begin{aligned} L_{1n} \frac{\Delta \theta \Delta \lambda}{\Delta \theta, \Delta \lambda} \frac{g_{nn,pn}^{\alpha \delta}}{4\pi} &= L_{1n} \sum_{\Delta \theta, \Delta \lambda \rightarrow 0}^{N-1} \sum_{j=0}^{N-1} \bar{Y}_{nn}^{\alpha}(\theta_1, \lambda_j) \bar{Y}_{pn}^{\alpha}(\theta_1, \lambda_j) \sin \theta_1 \frac{\Delta \theta \Delta \lambda}{4\pi} \\ &= \frac{1}{4\pi} \int_{\sigma} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) \bar{Y}_{pn}^{\alpha}(\theta, \lambda) d\sigma = \begin{cases} 0 & \text{if } n \neq p \\ 1 & \text{if } n = p \end{cases} \end{aligned}$$

because of the orthogonality relationships. In general, the variances of the noise are not going to follow a sinusoidal law, but one may reasonably expect (at least with more or less homogeneous noise) that the stability of the normal equations will not deteriorate with $\Delta \theta, \Delta \lambda \rightarrow 0$.

2.16 Global Adjustment and Collocation with Scattered Data

The efficient set up and inversion of the covariance matrix $(C_{zz} + D)$, or of the normal matrix G , depend on the regular nature of the grid. If not all nodes or blocks in the grid have data associated with them, the data is said to be scattered. The blanks or "holes" in the grid destroy the orderly structure of the matrices, making the application of the techniques previously discussed impossible. Yet so strong is this structure that, even in fragments, still it can be dealt with more efficiently than in the case of ordinary matrices of the same size.

(a) Full Region Bound by Lines of Latitude and Longitude

In the case when there is data at every point or block inside a "square" region limited by parallels and meridians, the partitioning of the data vector along the arcs or parallels inside the zone reveals a strong structure in the $(C_{zz} + D)$ matrix, if all the other assumptions made in paragraph (2.10) still apply.

If N_r is the number of rows and N_c the number of meridians that cross the region, then the covariance matrix will consist of N_r^2 blocks C^{ip} of dimension N_c , both persymmetrical and Toeplitz, though not circulant (i.e., the relationship $c_{jn}^{ip} = c_{j+1, n+1}^{ip}$ is fulfilled, but not $c_{0q}^{ip} = c_{N_c, q-1}^{ip}$); moreover the first row in each block does not have the property that $c_{0q}^{ip} = c_{0, N_c-q}^{ip}$. Clearly, though weaker than in the case of a global grid, there is a definite structure here that can be exploited to make both setting up and inverting the matrix more efficient.

Because each block C^{ip} is Toeplitz, only the N_c elements in its first row have to be computed, or about $\frac{1}{2} N_c N_r^2$ for the matrix as a whole, instead of $\frac{1}{2} N_c^2 N_r^2$; this amounts to a reduction in operations by a factor of N_c .

The solution of the equation $\hat{\underline{f}}_{nm}^{\alpha} = (C_{zz} + D)^{-1} \underline{c}_n^{\alpha} = \alpha_z$, for the optimal estimator vector $\hat{\underline{f}}_{nm}^{\alpha}$ for $\underline{C}_{nm}^{\alpha}$, can be obtained by a technique such as conjugate gradients, or similar, in which a finite number K of matrix-vector multiplications $(C_{zz} + D) \underline{v}_i$ (where $\underline{v}_0, \underline{v}_1, \dots, \underline{v}_i, \dots, \underline{v}_K$ are K intermediate N_p - vectors created during the solution) constitute the bulk of the computing effort. There is no need to go into the details of any specific technique, as the reader will find excellent descriptions in the literature (Householder, 1964, Luenberger, 1969). A discussion of the matrix-vector operation is sufficient here.

Let \underline{m}^i be the N_c - vector partition of the $N_c N_r$ data vector \underline{m} , containing the measurements along the i th parallel in the region, and let \underline{v}_i^i be the corresponding partition in any of the \underline{v}_i vectors. The product $(C_{zz} + D) \underline{v}_i^i = \underline{p}_i$ is, under such partition,

$$\underline{p}_i = [\underline{p}_i^0 \dots \underline{p}_i^{N_r-1}]^T$$

with

$$\underline{p}_i^i = \sum_{p=0}^{N_r-1} C^{ip} \underline{v}_i^p \quad (2.90)$$

so the whole matrix-vector multiplication can be broken up into N_r^2 products $C^{ip} \underline{v}_i^p = \underline{h}_i^p$. Because C^{ip} is a Toeplitz matrix, the N_c components of \underline{h}_i^p can be obtained by "weighted running averages" or discrete convolution of the elements of \underline{v}_i^p with those in the first row of C^{ip} . Such convolution can be calculated efficiently using the Fast Fourier Transform algorithm (see, for instance, Brigham, 1974, Ch. 13). Therefore, all N_r^2 products involve $O(N_c N_r^2)$ operations, and since there are K matrix-vector multiplications in the whole procedure, the total number of operations needed to obtain $\hat{\underline{f}}_{nm}^{\alpha}$ amounts to $O(K N_c N_r^2)$. For conjugate gradients, K does not exceed (in theory) $N_p = N_r N_c$, so there should be $O(N_c^2 N_r^3)$ operations altogether. If $(C_{zz} + D)$ were handled by conventional techniques, disregarding its well defined structure, the number would be $O(N_c^3 N_r^3)$, so the increase in efficiency is $O(N_c)$, the same as for the setting up.

(b) Arbitrarily Scattered Data

It is common in geodesy and in geophysics to have a set of measurements scattered throughout the globe, without the data being on the nodes of a regular grid or without all σ_i^2 , being equal along parallels (nonhomogenous noise). If the set is dense enough, however, it is possible to interpolate the data quite reliably on the closest nodes of a conveniently chosen grid. Assuming that this is done, and that the accuracies of the interpolated values are known well

enough, then the problem can be dealt with by conventional least squares or by collocation. In general, there will be blanks or "holes" irregularly distributed over the sphere, the actual data points falling among them in no precise pattern. This problem will be considered here as a least squares adjustment problem. If the data has little or no power above the Nyquist frequency of the grid on which it has been interpolated, then the extension of the ideas that follow to collocation is quite simple, according to paragraph (2.13).

Consider the element $g_{\alpha\beta}^{\alpha\beta}$ of the matrix $G = A^T D^{-1} A$

$$g_{\alpha\beta}^{\alpha\beta} = \sum_{i=0}^{N-1} \bar{P}_{\alpha i} (\cos \theta_i) \bar{P}_{\beta i} (\cos \theta_i) \sum_{j=0}^{N-1} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} p j \Delta \lambda W_{ij} \sigma_{ij}^2 \quad (2.91)$$

where $W_{ij} = 0$ if the point ij is blank, otherwise $W_{ij} = 1$; σ_{ij}^2 is the variance of the noise n_{ij} . In general σ_{ij}^2 is a function both of i and of j . Clearly, matrix D is taken to be diagonal (i.e., uncorrelated noise).

From the relationships

$$\begin{aligned} \cos m j \Delta \lambda \cos p j \Delta \lambda &= \frac{1}{2} [\cos (m+p) j \Delta \lambda + \cos (m-p) j \Delta \lambda] \\ \cos m j \Delta \lambda \sin p j \Delta \lambda &= \frac{1}{2} [\sin (m+p) j \Delta \lambda - \sin (m-p) j \Delta \lambda] \\ \sin m j \Delta \lambda \sin p j \Delta \lambda &= -\frac{1}{2} [\cos (m+p) j \Delta \lambda - \cos (m-p) j \Delta \lambda] \end{aligned}$$

and calling

$$C_r^{\alpha} \equiv \begin{cases} C_r^{\alpha 0} = \frac{1}{2} \sum_{j=0}^{N-1} \cos r j \Delta \lambda \sigma_{ij}^2 W_{ij} \\ C_r^{\alpha 1} = \frac{1}{2} \sum_{j=0}^{N-1} \sin r j \Delta \lambda \sigma_{ij}^2 W_{ij} \end{cases} \quad (2.92)$$

where $-N < r < 2N$, follows

$$\sum_{j=0}^{N-1} \sigma_{ij}^2 W_{ij} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m j \Delta \lambda \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} p j \Delta \lambda = [C_{(-r)}^{\alpha} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} p j \Delta \lambda + (-1)^{\alpha} C_{(r)}^{\alpha} \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} p j \Delta \lambda] \quad (2.93)$$

Moreover, because

$$\begin{aligned} \cos (-r) j \Delta \lambda &= \cos r j \Delta \lambda \\ \sin (-r) j \Delta \lambda &= -\sin r j \Delta \lambda \\ \cos r j \Delta \lambda &= \cos (2N - r) j \Delta \lambda \\ \sin r j \Delta \lambda &= -\sin (2N - r) j \Delta \lambda \end{aligned} \quad (2N \Delta \lambda = 2\pi)$$

follows

$$\begin{aligned} C_{-r}^{\alpha} &= C_r^{\alpha} (-1)^{\alpha} \\ C_{2N-r}^{\alpha} &= C_r^{\alpha} (-1)^{\alpha} \end{aligned}$$

where the last one is of special interest when $r \geq N$. So only the $C_r^i \alpha$ with $0 < r < N$ are needed. Finally, calling

$$g_{n\alpha, p\beta}^i = \bar{P}_{n\alpha}(\cos \theta_i) \bar{P}_{p\beta}(\cos \theta_i) [C_{n-\alpha}^i |\alpha-\beta| + (-1)^{2\alpha+\beta} C_{(n+\alpha)}^i |\alpha-\beta|] \quad (2.94)$$

(where $|\alpha-\beta|$ is the absolute value of $\alpha-\beta$)

it is

$$g_{n\alpha, p\beta}^{\alpha, \beta} = \sum_{i=0}^{N-1} g_{n\alpha, p\beta}^i \quad (2.95)$$

Assume that out of N rows the grid has only I with any data in them, so at least one $W_{i,j} \neq 0$ in each. Once the corresponding $C_r^i \alpha$ have been obtained by computing the discrete Fourier transform of $\sigma_{i,j}^{-2} W_{i,j}$ along each row with data ($O(N^3)$ operations for the whole grid), what remains is to get the I non-zero terms $g_{n\alpha, p\beta}^i$ that form each element $g_{n\alpha, p\beta}^{\alpha, \beta}$ of G . As there are about $\frac{1}{2}N^4$ such elements that are different, and I is $O(N)$ (except for very sparse data sets), the total number of operations needed to create G is $O(N^3) + O(N^5)$, or virtually $O(N^5)$. If the $g_{n\alpha, p\beta}^{\alpha, \beta}$ were computed according to (2.91) as it is written, instead of according to its reduced version. (2.95), the number of operations would be $O(2N^2 (N)^4)$, or $O(N^6)$, so the gain in efficiency allowed by this approach is $O(N)$, which is the same as in the case studied in the first part of this paragraph, where the data completely filled a "square" sector of the sphere. This count does not include the time needed to obtain the $\bar{P}_{n\alpha}(\cos \theta_i)$, as these can be pre-computed once and kept on disk or tape for repeated use.

The number of operations can be reduced further, almost by half, by taking advantage of the fact that $\bar{P}_{n\alpha}(\cos \theta_i)$ is a common factor in all $g_{n\alpha, p\beta}^i$ with the same n and the same m . Furthermore, if there are pairs of rows i and $N-i-1$ (i.e., symmetrical with respect to the Equator) where both rows contain some data, then

$$g_{n\alpha, p\beta}^i + g_{n\alpha, p\beta}^{N-i-1} = \bar{P}_{n\alpha}(\cos \theta_i) \bar{P}_{p\beta}(\cos \theta_i) [(C_{n-\alpha}^i |\alpha-\beta| + (-1)^{2\alpha+\beta} C_{n+\alpha}^{N-i-1} |\alpha-\beta|) + (-1)^{2\alpha+\beta} (C_{n-\alpha}^{N-i-1} |\alpha-\beta| + (-1)^{2\alpha+\beta} C_{n+\alpha}^i |\alpha-\beta|)] \quad (2.96)$$

which leads to further savings in computing at the cost of additional programming complexity. These economies are important, but they will not bring the number of operations much below $O(N^5)$ unless the grid is so sparse that I is much smaller than N .

Notice that the normal matrix $G = A^T D^{-1} A$ is created here without actually forming the observation equations matrix A . This means considerable savings in computing and in storage requirements. As for the right hand side of the normals $A^T D^{-1} \underline{m} = \underline{b}$, the elements of \underline{b} are given by the formula

$$b_k = \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \bar{P}_{n,n}(\cos \theta_i) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda \sigma_{ij}^{-2} W_{ij} m_{ij} \quad (2.97)$$

where $k = n^2 + \alpha n + m + 1$ and $W_{ij} = 0$ if there is no data at the point ij , as before. Such expression is of the quadratures type, and can be computed efficiently by the corresponding algorithm of section 1, also without first creating A . Finally, the residuals vector $\underline{v} = \underline{m} - A \hat{\underline{c}}$, usually of interest, can be obtained as the difference between the data \underline{m} and the values of

$$\hat{z}(\theta_i, \lambda_j) = \sum_{\alpha=0}^1 \sum_{n=0}^1 \sum_{m=0}^n \hat{C}_{n,n}^{\alpha} \bar{P}_{n,n}(\cos \theta_i) \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} m_j \Delta \lambda$$

computed by means of the appropriate synthesis procedure given in section 1. Thus \underline{v} can be found without knowing A explicitly.

The normal equations can be solved by means of conjugate gradients or a similar method involving M matrix-vector products $\underline{x}_t = G \underline{h}_t$, where \underline{h}_t is part of a sequence of intermediate vectors $\underline{h}_0, \underline{h}_1, \dots, \underline{h}_t, \dots, \underline{h}_M$.

Introducing the notation

$$\underline{h}_{n,n,t}^{\alpha} = \underline{h}_{kt}$$

where $k = n^2 + \alpha n + m + 1$, and calling G^i to the matrix of all $g_{n,n,pq}^{i,\alpha,\beta}$,

$$\text{so } G = \sum_{i=0}^{N-1} G^i \text{ and } \underline{x}_t = \sum_{i=0}^{N-1} G^i \underline{h}_t = \sum_{i=0}^{N-1} \underline{x}_t^i \quad (2.98)$$

then each element $x_{n,n,t}^{\alpha,i}$ of the product vector \underline{x}_t^i is of the form

$$x_{n,n,t}^{\alpha,i} = \sum_{p=0}^{N-1} \sum_{q=0}^p \sum_{\beta=0}^q g_{n,n,pq}^{i,\alpha,\beta} h_{pq,t}^{\beta} = \bar{P}_{n,n}(\cos \theta_i) \sum_{pq\beta} [C_{n+q}^{i|\alpha-\beta|} + (-1)^{2\alpha+\beta} C_{n-q}^{i|\alpha-\beta|}] \cdot \bar{P}_{pq}(\cos \theta_i) h_{pq,t}^{\beta} \quad (2.99)$$

Every $h_{pq,t}^{\beta}$ above multiplies always the same $\bar{P}_{pq}(\cos \theta_i)$; calling

$$d_{pq}^{i\beta} = \bar{P}_{pq}(\cos \theta_i) h_{pq,t}^{\beta}$$

expression (2.99) becomes

$$x_{n,n,t}^{\alpha,i} = \bar{P}_{n,n}(\cos \theta_i) \sum_{pq\beta} [C_{n+q}^{i|\alpha-\beta|} + (-1)^{2\alpha+\beta} C_{n-q}^{i|\alpha-\beta|}] d_{pq}^{i\beta}$$

All quantities inside the square brackets are constant if α, β, m , and q are the same. Grouping equal factors together

$$\begin{aligned} x_{n,n,t}^{\alpha,i} &= \bar{P}_{n,n}(\cos \theta_i) \sum_{q=n}^{N-1} \sum_{\beta=0}^q [C_{n+q}^{i|\alpha-\beta|} + (-1)^{2\alpha+\beta} C_{n-q}^{i|\alpha-\beta|}] \sum_{p=0}^{N-1} d_{pq}^{i\beta} \\ &= \bar{P}_{n,n}(\cos \theta_i) \sum_{q=n}^{N-1} \sum_{\beta=0}^q [C_{n+q}^{i|\alpha-\beta|} + (-1)^{2\alpha+\beta} C_{n-q}^{i|\alpha-\beta|}] D_q^{i\beta} \end{aligned} \quad (2.100)$$

$$\text{where } D_q^{i\beta} = \sum_{p=0}^{N-1} d_{pq}^{i\beta}$$

There are (N^2) products needed to form all d_i^{β} ; (N^3) sums to compute the $2N \cdot D_i^{\beta}$, and (N^2) further multiplications by the $P_{nm}(\cos \theta_i)$ to form all $x_n^{\alpha\beta}$. As there are I non-zero G^i , there are, per matrix-vector product, $I \times O(N^2)$ operations, or $O(N^3)$ if the data is not too sparse. M , the number of matrix-vector products in the solution, is $O(N^2)$, so the total comes to $O(N^5)$. Inverting G by the usual methods and without having regard for its structure involves $O(N^3)$ operations. The gain in efficiency is, again, $O(N)$.

Besides providing a convenient way of demonstrating how the properties of G can be exploited to make its inversion more efficient (or the solution of the normal equations, both approaches are equivalent), conjugate gradients is interesting on its own right. Sparse data sets with a poor distribution will result in ill-conditioned normals, so the inversion of G may be numerically impossible. So-called iterative methods, such as conjugate gradients, usually improve the initial guess (represented by h_0) of the correct values of the unknowns, at least for the first few iterations. Improvement here means a reduction in the quadratic form being minimized, such as the mean square value of the residuals. If the initial guess is a good one, and present day spherical harmonic coefficients of the gravity field are reasonable good for degrees up to 30 or so, then a few iterations are likely to improve this guess, and to produce reasonable estimates of those coefficients that, being wholly unknown, are taken to be zero at the start. If a "few" iterations are much less than the maximum N^2 , then a reduction in computing time of $O(N^2)$ takes place. This might allow scientists to "extend" existing models to much higher degree and order than at present, simply by obtaining approximate solutions of this type.

Clearly, all that has been said here regarding G and least squares adjustment applies equally well to $(G + C^{-1})$ and least squares collocation. Though the formulas have been developed on the basis of a point values' formulation, their extension to area means is not difficult.

While the descriptions of the methods for estimating the \bar{C}_{nn}^α discussed here have been confined to the case where all the data are of one kind (i.e., gravity anomalies only, or magnetic anomalies only, or geopotential numbers, etc.) their extension to mixed data sets is immediate, provided that all values are globally distributed according to grids of the same $\Delta\lambda$, although the latitudes need not be the same as well.

Note on the Accuracies of the \hat{C}_{nn}^α when Using Conjugate Gradients:

Besides the estimates of the coefficients, one usually wants to know the accuracies of those estimates. When a few iterations of conjugate gradients are used to update some initial estimates in the efficient way described above, what are the accuracies of the improved coefficients? In the course of the conjugate gradients procedure (see references), the conjugate directions \underline{v}_k of $G = (A^T D^{-1} A + C^{-1} + Q_s^{-1})$ are generated (Q_s^{-1} is the variance covariance matrix of the initial estimates, as explained in paragraph (2.18) part (b)), together with the scalars $\underline{v}_k^T G \underline{v}_k = \alpha_k$. The conjugate directions have the property

$$\underline{v}_k^T G \underline{v}_p = 0$$

for $k \neq p$. The estimator implied by R iterations of this procedure is

$$\hat{\underline{c}} = \tilde{G} (A^T D^{-1} \underline{m} + Q_s^{-1} \underline{c}_s) = \tilde{G} A^T D^{-1} (\underline{z} + \underline{n}) + \tilde{G} Q_s^{-1} (\underline{c} + \Delta \underline{c}_s)$$

where \underline{c}_s is the vector of initial estimates, $\Delta \underline{c}_s$ the errors in this vector, and

$$\tilde{G} = \sum_{k=1}^R \alpha_k^{-1} \underline{v}_k \underline{v}_k^T$$

The variance-covariance matrix of the updated errors, for ordinary least squares, is (assuming that $E\{\underline{n} \Delta \underline{c}_s^T\} = 0$)

$$\begin{aligned} E\{(\tilde{G} A^T D^{-1} \underline{n})(\tilde{G} A^T D^{-1} \underline{n})^T\} + E\{(\tilde{G} Q_s^{-1} \Delta \underline{c}_s)(\tilde{G} Q_s^{-1} \Delta \underline{c}_s)^T\} &= \tilde{G} A^T D^{-1} D D^{-1} A \tilde{G} \\ + \tilde{G} Q_s^{-1} Q_s Q_s^{-1} \tilde{G} &= \tilde{G} (A^T D^{-1} A + Q_s^{-1}) \tilde{G} = \tilde{G} G \tilde{G} \\ &= \sum_{k=1}^R \alpha_k^{-1} \underline{v}_k \alpha_k \underline{v}_k^T \alpha_k^{-1} = \sum_{k=1}^R \alpha_k^{-1} \underline{v}_k \underline{v}_k^T = \tilde{G} \end{aligned}$$

Therefore, the variances of the errors in the \hat{C}_{nn}^α are equal to the corresponding diagonal elements of \tilde{G} :

$$\sigma_{\eta_{nn}^\alpha}^2 = \sum_{k=1}^R \alpha_k^{-1} v_{k,nn}^2 \alpha_k$$

where $v_{k,nn} \alpha_k$ is the element of \underline{v} corresponding to \hat{C}_{nn}^α in $\hat{\underline{c}}$ (\underline{v} and $\hat{\underline{c}}$ have the same dimension). The same result applies to least squares collocation. All that is needed to obtain the $\sigma_{\eta_{nn}^\alpha}^2$, according to the formula

above, is knowledge of the α_k and of the y_k , which can be saved as they are created, during the k th iteration of the procedure. For the efficient algorithm to be applicable, Q_s^{-1} must have a suitable structure. One case is when Q_s^{-1} is diagonal, or can be satisfactorily approximated by a diagonal matrix.

2.17 The Error Matrix in the Band-Limited Case

From (2.41) and (2.77) results

$$\begin{aligned} E_T &= C - C_{oz} (C_{zz} + D)^{-1} C_{oz}^T = C - F C_{oz}^T = C - (A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1} A C \\ &= [I - (A^T D^{-1} A + C^{-1})^{-1} A^T D^{-1} A] C = (A^T D^{-1} A + C^{-1})^{-1} [(A^T D^{-1} A + C^{-1}) - A^T D^{-1} A] C \\ &= (A^T D^{-1} A + C^{-1})^{-1} \end{aligned} \quad (2.101)$$

In the case of the best unbiased estimator C^{-1} is not present in the normal matrix, so (2.101) becomes

$$E_T = (A^T D^{-1} A)^{-1} = G^{-1} \quad (2.102)$$

which is the well known expression of the error matrix for ordinary least squares.

Because of the block-diagonal structure of the variance-covariance matrix,⁽¹⁾ the estimates of \bar{C}_{nn}^α of different orders are uncorrelated. The diagonal elements of the variance-covariance matrix are the variances of the estimated coefficients total errors (i.e., sampling plus propagated noise), in (2.101).

Obtaining the variances of the \bar{C}_{nn}^α in band-limited collocation is formally identical to getting the variances of the estimates in ordinary least squares, according to (2.101) and to (2.102)

2.18 The Use of a priori Information on the Coefficients

Assume that all coefficients up to some degree and order M are approximately known, and that Q_s is the variance-covariance matrix of their errors. This could be the case where a model of the gravity field has been obtained from data gathered using artificial satellites, complete to degree M , and terrestrial data is to be used to improve the existing coefficients and obtain new ones beyond degree M .

Three possible approaches to this question will be discussed here, using a "point values" formulation in the first two cases for simplicity.

¹ When the data set is not sparse.

(a) Simple weighted averages:

The terrestrial data can be used separately to obtain a model to degree and order $M_T > M$, together with the variance-covariance matrix Q_T of the coefficients' errors. To combine the satellite and the terrestrial coefficients one can set up the following observation equations

$$\begin{bmatrix} \underline{c}_s \\ \underline{c}_T \end{bmatrix} + \begin{bmatrix} \underline{s} \\ \underline{t} \end{bmatrix} = \begin{bmatrix} I_s \\ I_T \end{bmatrix} \underline{c} \quad (2.103)$$

where \underline{c} is the vector potential coefficients and I_s is the $(M+1)^2 \times (M+1)^2$ unit matrix augmented with zeroes on the right, and I_T is the $(M_T+1)^2 \times (M_T+1)^2$ unit matrix, while \underline{c}_s is the vector of 'satellite' coefficients and \underline{c}_T the vector of 'terrestrial' coefficients; \underline{s} and \underline{t} are the corresponding vectors of residuals. The best linear unbiased estimator for the combined system of observations is

$$\begin{aligned} \hat{\underline{c}} &= ([I_s \ I_T] \begin{bmatrix} Q_s^{-1} & 0 \\ 0 & Q_T^{-1} \end{bmatrix} \begin{bmatrix} I_s \\ I_T \end{bmatrix})^{-1} [I_s \ I_T] \begin{bmatrix} Q_s^{-1} & 0 \\ 0 & Q_T^{-1} \end{bmatrix} \begin{bmatrix} \underline{c}_s \\ \underline{c}_T \end{bmatrix} \\ &= (Q_s^{-1} + Q_T^{-1})^{-1} (Q_s^{-1} \underline{c}_s + Q_T^{-1} \underline{c}_T) \end{aligned} \quad (2.104)$$

If Q_s and Q_T are the inverses of ordinary least squares normal matrices G_s and G_T , then the error matrix of the combined solution

$$E = (Q_s^{-1} + Q_T^{-1})^{-1} = (G_s + G_T)^{-1}$$

corresponds to the propagated noise only. If they are "collocation" matrices of the type $(A^T D^{-1} A + C^{-1})^{-1}$ (see expression (2.77)), then the error matrix includes the effect of the sampling error as well. Most satellite models are, to date, "least squares-type" and it would be incongruous to combine them with "collocation-type" models, terrestrial or otherwise. The problem need not be a serious one, because geodetic spacecrafts so far have orbited at altitudes of 800 km or more, where the field is much smoother than at the surface, so the sampling errors are bound to be small compared to the propagated data errors reflected by Q_s .

In the case where the terrestrial model has been derived from a regularly sampled data set using the "band-limited approach" of previous paragraphs, Q_T is block-diagonal, and those blocks corresponding to orders $m > M$ are identical to the corresponding blocks in $(Q_s^{-1} + Q_T^{-1})^{-1}$. From this it is not difficult to conclude that the coefficients in the combined model up to order M will be somewhat different (and presumably better) than those in either the satellite or the terrestrial sets, while those above M will be identical to the corresponding terrestrial coefficients.

(b) A priori values included as data in the adjustment:

Consider the system of observation equations

$$\begin{bmatrix} \underline{m} \\ \underline{c}_s \end{bmatrix} + \begin{bmatrix} \underline{v} \\ \underline{s} \end{bmatrix} = \begin{bmatrix} \underline{A} \\ \underline{I}_s \end{bmatrix} \underline{c} \quad (2.105)$$

which is the system (2.66) augmented with equations of the type

$$\hat{C}_{ns}^{\alpha} - e_{ns}^{\alpha}(s) = \bar{C}_{ns}^{\alpha}$$

where $e_{ns}^{\alpha}(s)$ is the error in the "satellite model" coefficient $\hat{C}_{ns}^{\alpha}(s)$. The normal matrix of collocation is

$$E_r^{-1} = ([\underline{A}^T \underline{I}_s^T] \begin{bmatrix} \underline{D}^{-1} & \underline{O} \\ \underline{O} & \underline{Q}_s^{-1} \end{bmatrix} [\underline{A} \\ \underline{I}_s] + \underline{C}^{-1}) \quad (2.106)$$

and the optimal estimator of the band-limited type is

$$\hat{\underline{c}} = (\underline{A}^T \underline{D}^{-1} \underline{A} + \underline{Q}_s^{-1} + \underline{C}^{-1})^{-1} (\underline{A}^T \underline{D}^{-1} \underline{m} + \underline{Q}_s^{-1} \underline{c}_s) \quad (2.107)$$

Once more, if the data in \underline{m} has been sampled regularly on the sphere, the estimated coefficients will be affected by the existence of a priori values only if their order is no higher than M . Naturally, this is a desirable situation. Also it is important that the error measures corresponding to E_r and \underline{Q}_s be congruous, though this is probably not very important in the case of satellite models obtained with high-orbiting spacecraft. Notice that (a) and (b) are equivalent when $\underline{Q}_r = (\underline{A}^T \underline{D}^{-1} \underline{A})^{-1}$ and when \underline{C}^{-1} is excluded from (2.106)-(2.107). In other words: these first two approaches are equivalent for ordinary least squares.

(c) The method of Kaula and Rapp:

W. Kaula (1966) proposed a technique for simultaneously filtering errors out of a terrestrial data set and improving the coefficients of a satellite model. This method was later developed by R. Rapp (1968), who more recently (1978) used it to improve a global data set of mean $1^\circ \times 1^\circ$ anomalies by combining it with the potential coefficients of the GEM-9 model. This adjusted data set was used by the author of this report to create the $5^\circ \times 5^\circ$ mean anomalies analysed in one of the numerical experiments of section 3.

The idea is to satisfy condition equations of the type

$$\hat{C}_{ns}^{\alpha} + d_{ns}^{\alpha} - (4\pi\gamma(n-1)\beta_n S^{n+2})^{-1} \sum_{i=0}^{n-1} \sum_{j=0}^{2n-1} \int_{\sigma_{1j}} \bar{Y}_{ns}^{\alpha}(\theta, \lambda) d\sigma (\bar{g}_{1j} + v_{1j}) = 0 \quad (2.108)$$

while minimizing the quadratic form

$$g(\underline{d}, \underline{v}) = \underline{d}^T \underline{Q}^{-1} \underline{d} + \underline{v}^T \underline{D}^{-1} \underline{v} \quad (2.109)$$

where \underline{v} is the vector of corrections v_i to the mean gravity anomalies Δg_{ij} , and \underline{d} is the vector of corrections d_{nn}^α to the satellite potential coefficients $\hat{C}_{nn}^\alpha(\gamma)$; β_n is the n th degree Pellinen factor discussed in paragraph (4.3), γ is the mean value of equatorial gravity, and S is the ratio between the radius of the Earth's largest inner geocentric sphere and the mean Earth radius. In matrix form, the condition equations (2.108) are

$$\underline{c}_s + \underline{d} - \underline{A}^T \underline{\Delta g} - \underline{A}^T \underline{v} = 0 \quad (2.110)$$

where \underline{A} is a $2N^2 \times (M+1)^2$ matrix having columns $\underline{a}_{nn}^\alpha$ of the form

$$\underline{a}_{nn}^\alpha = (4\pi(n-1)\gamma S^{n+2} \beta_n)^{-1} \left[\int_{\sigma_{00}} \bar{Y}_{nn}^\alpha(\theta, \lambda) d\sigma \dots \int_{\sigma_{N-1, 2N-1}} \bar{Y}_{nn}^\alpha(\theta, \lambda) d\sigma \right]^T$$

so, except for the factor $(4\pi(n-1)\gamma S^{n+2} \beta_n)^{-1}$, \underline{A} is the "area means version" of the matrix of system (2.66), and has the same properties as the "A" matrices considered so far.

The optimal estimates are given by the expressions

$$\hat{\underline{c}} = \underline{c}_s + \underline{d} \quad (2.111-a)$$

$$\hat{\underline{\Delta g}} = \underline{\Delta g} + \underline{v} \quad (2.111-b)$$

where

$$\underline{d} = -((\underline{A}^T \underline{D} \underline{A})^{-1} + \underline{Q}_s^{-1})^{-1} (\underline{A}^T \underline{D} \underline{A})^{-1} (\underline{c}_s - \underline{A}^T \underline{\Delta g}) \quad (2.112)$$

and

$$\underline{v} = \underline{D} \underline{A} \underline{Q}_s \underline{d} \quad (2.113)$$

If the data set is both complete and of uniform quality, the matrix $\underline{A}^T \underline{D} \underline{A}$ has the block structure first discussed in paragraph (2.15). The presence of \underline{D} instead of \underline{D}^{-1} makes no difference to the calculations needed to set up and invert the matrix: the procedures are those already explained. Terrestrial data, however, is usually both scattered and of varying quality (i.e., different noise variances). For this type of data, therefore, the methods for scattered measurements given in paragraph (2.16) could be used.

2.19 Optimal Estimation over a Band of Spatial Frequencies

Assume that the signal is of the type

$$\underline{m} = \underline{A} \underline{c} + \underline{n}$$

where A and \underline{c} may now be infinite (i.e., all degrees from 0 to ∞ may be present). If \underline{c}' is a sub-vector of \underline{c} comprising, say, the first N^2 coefficients, and if $\underline{\hat{s}}$ is a vector of estimates of a function s at a given set of points on the sphere, such that the values of s depend only on those of \underline{c}' according to the relationship

$$\underline{s} = B \underline{c}' \quad (2.114)$$

where B is some matrix of appropriate size, not necessarily of the same type as A , then the optimal estimator for $\underline{\hat{s}}$ is, according to (2.39), (2.75),

$$\begin{aligned} \underline{\hat{s}} &= C_{sz} (C_{zz} + D)^{-1} \underline{m} \\ &= M \left\{ \underline{s} \underline{z}' \right\} (C_{zz} + D)^{-1} \underline{m} = M \left\{ B \underline{c}' \underline{c}'^T A^T (C_{zz} + D)^{-1} \right\} \\ &= B C' A^T (C_{zz} + D)^{-1} \underline{m} \quad (\text{where } C' = M \left\{ \underline{c}' \underline{c}'^T \right\}) \\ &= B C_{s'z} (C_{zz} + D)^{-1} \underline{m} \end{aligned}$$

or

$$\underline{\hat{s}} = B \underline{\hat{c}}' \quad (2.115)$$

according to (2.40).

Expression (2.115) indicates that the optimal estimates of a band-limited function s from data \underline{m} are identical to the values of s obtained from the optimal estimates of the coefficients \underline{c}' by means of the relationship $\underline{\hat{s}} = B \underline{\hat{c}}'$.

3. Numerical Examples

This section presents several computed examples to illustrate some of the ideas and methods discussed earlier on. The question posed here is, basically, that of the accuracy of the various procedures, and is answered by means of error analysis carried out with the formulas for error variance developed in section 2, and also by analysing simulated data and comparing the recovered coefficients to the original ones in order to find the actual errors. A comparison of the rms of these errors with the theoretical rms (i.e., the square root of the variance) provides both a check on each set of results and, more important, shows just how adequate an error measure the theoretical rms can be. Besides error analysis and simulations, this section shows, in the last paragraph, the results of the harmonic analysis of a real data set: a $5^\circ \times 5^\circ$ equal angular set of mean gravity anomalies covering the whole Earth, from which the coefficients of the disturbing potential have been recovered to degree and order 36 by means of least squares collocation, using the "Toeplitz matrix" approach of paragraphs (2.10) and (2.11).

3.1 Generation and Analysis of Simulated Data

As explained in the preceding section, the variance of the error in the estimate of \bar{C}_{nn}^α depends on the power spectrum (or covariance function) of the signal, and on the variance-covariance matrix of the noise. The propagation of the noise is quite straightforward, and anybody who has had any practical experience with adjustments of geodetic networks and the like already has enough "feeling" for this part of the error measure, and is capable of understanding its significance when its value is given to him. The part corresponding to the sampling error is somewhat different; it involves a rather unusual geometric average over rotations, and this type of error measure, while not exactly new (collocation, based on this measure, has been around since the mid-sixties) is not so familiar to geodesists yet, and its use, in harmonic analysis in particular, far from common practice. For this reason, it is probably fair to the readers to provide some illustration of how "close" this part of the error measure is to the actual sampling error that occurs when data of the assumed power spectrum is analyzed in any of the ways discussed so far to recover spherical harmonic coefficients. By "close" one means that the actual numbers measuring the theoretical and the actual variances (or rms) should differ from each other by a small percentage, or some equally clear-cut criterion.

The theoretical variance considered here is the variance of the estimation errors per degree $\sigma_{\epsilon_n}^2$, defined in terms of the error measure of section 2 as follows (see paragraph (2.8))

$$\sigma_{\epsilon_n}^2 = \sum_{n=0}^{\infty} \sum_{\alpha=0}^1 \sigma_{n\alpha}^2 = \sum_{n=0}^{\infty} \sum_{\alpha=0}^1 \left(\frac{\sigma_n^2}{(2n+1)} - 2 \underline{c}_{n\alpha}^T \alpha_z \underline{f}_{n\alpha}^\alpha + (\underline{f}_{n\alpha}^\alpha)^T (C_{zz} + D) \underline{f}_{n\alpha}^\alpha \right) \quad (3.1)$$

The rms of the error is the square root of this variance, and the ratio of this rms to the rms per coefficient:

$$\frac{\sigma_{\epsilon_n}}{\sigma_n} = \left(1 - \sigma_n^{-2} \sum_{n=0}^{\infty} \sum_{\alpha=0}^1 (2 \underline{c}_{n\alpha}^T \alpha_z \underline{f}_{n\alpha}^\alpha - (\underline{f}_{n\alpha}^\alpha)^T (C_{zz} + D) \underline{f}_{n\alpha}^\alpha) \right)^{\frac{1}{2}} \quad (3.2)$$

multiplied by 100, or the percentage rms error per degree, is the theoretical quantity to be compared to the "actual" percentage rms error per degree derived from the analysis of simulated data with the same statistical characteristics (i.e., σ_n^2 , C_{zz} , and D) in formula (3.2).

The $\sigma_{\epsilon_n}^2$ were computed with subroutine NORMAX (Appendix B).

To obtain the actual percentage rms error per degree, sets of simulated data were created on full regular grids as follows: the artificial data consisted of area means computed globally, using the algorithm outlined in paragraph (1.7) and subroutine SSYNTH (Appendix B), on the basis of expression (1.2). The $\bar{C}_{n\alpha}^\alpha$, complete to degree and order $N_{max} \gg N = \frac{2\pi}{\Delta\lambda}$, came from sequences of random numbers. The random numbers, obtained using the IMSL subroutine "GGNOR" with generating "seeds" of the order of 10^4 , were scaled to give them the desired degree variances σ_n^2 . For each simulation, a sequence of $(N_{max} + 1)^2$ numbers was obtained, the first corresponding to \bar{C}_{00} , the second, third, and fourth to \bar{C}_{10} , \bar{C}_{11} , and \bar{S}_{11} , respectively, and so forth. If r_{n0} , r_{n1}^α , \dots , $r_{n\alpha}^\alpha$ were the $(2n+1)$ numbers corresponding to degree n , then the scaling that resulted in the corresponding $\bar{C}_{n\alpha}^\alpha$ was

$$\bar{C}_{n\alpha}^\alpha = r_{n\alpha}^\alpha \left[\frac{\sigma_n}{\sqrt{\sum_{n=0}^{\infty} \sum_{\alpha=0}^1 (r_{n\alpha}^\alpha)^2}} \right] \quad (3.3)$$

The harmonic coefficients $\bar{C}_{n\alpha}^\alpha$ obtained in this way were the "actual" coefficients to which the $\hat{C}_{n\alpha}^\alpha$, recovered by some of the procedures described in section 2, were then compared to obtain the actual percentage rms errors per degree

$$\zeta_n = \left[\sum_{n=0}^{\infty} \sum_{\alpha=0}^1 (\hat{C}_{n\alpha}^\alpha - \bar{C}_{n\alpha}^\alpha)^2 \right]^{\frac{1}{2}} \sigma_n^{-1} \times 100 \quad (3.4)$$

The analysis of the simulated data was done with subroutine HARMIN (appendix B). This type of numerical experiment was carried out three or more times in each case, varying only the seed used to generate the random numbers, much as a Montecarlo-type of analysis is conducted. The seeds were chosen widely apart, to ensure that the correlation between "trials" would be virtually nil.

The maximum degree and order in the set of artificial coefficients, N_{max} , was chosen so that the power in the mean values above degree N_{max}

$$P_{N_{max}} = \sum_{n=N_{max}+1}^{2000} \beta_n^2 \sigma_n^2 \quad (3.5)$$

were less than 1% of the power between degrees 0 and 2000. The β_n are the Pellinen coefficients, discussed in paragraph (4.3), corresponding to $5^\circ \times 5^\circ$ area means. The values used for the degree variances σ_n^2 had been empirically obtained from terrestrial data in the manner described below. The data were supposed to be noise-free, as only the sampling part of the error was studied in this way, for the reasons given at the beginning of this paragraph. The estimators being linear, the propagated noise and the sampling error merely add arithmetically to each other, and can therefore be studied separately, if so desired. Summing up, it can be said that the simulated data consisted in global data sets of artificial gravity anomalies, averaged over equal angular grids.

The empirical degree variances were obtained as follows: up to degree 100 they were those implied by a set of coefficients, complete to degree 180, obtained by R. Rapp and associates at O.S.U. from a global data set of $1^\circ \times 1^\circ$ mean anomalies. Above degree 100, the $\sigma_n^2(\Delta g)$ were obtained from a model of the form

$$\sigma_n^2(\Delta g) = (n-1) \left(\frac{\alpha_1}{(n+A)} S_1^{n^2} + \frac{\alpha_2}{(n+B)} \frac{S_2^{n^2}}{(n-2)} \right) [\text{mgal}^2] \quad (3.6)$$

(Moritz, 1976), where the parameters α_1 , α_2 , S_1 , S_2 , A and B have been adjusted to fit existing gravimetric data, satellite altimetry, satellite field models, and other geophysical data. The parameters used in most examples were

$$\begin{array}{lll} \alpha_1 = 3.4050 & S_1 = 0.998006 & A = 1. \\ \alpha_2 = 140.03 & S_2 = 0.914232 & B = 2. \end{array}$$

corresponding to the best model of this type given in a report by R. Rapp (1979) who, in the same work, discusses also the empirical degree variances obtained from his 180, 180 field model. The degree variances implied by (3-6) with the parameters listed above are also very similar to those obtained by quite different means by Wagner and Colombo (1979), who analyzed the (Fourier) power spectrum of short arcs of GEOS-3 altimetry, and converted their average to a spherical harmonics spectrum using formulas that follow from the relationship between spherical harmonics and Fourier series. The empirical variances for $n \leq 100$ are included in the listing of subroutine NVAR, in Appendix B.

In order to understand how critical the choice of empirical degree variances is to the theoretical and actual errors, a different two-term model obtained by C. Jekeli (1978) was used as well. This model has the following parameter

values

$$\begin{array}{lll} \alpha_1 = 18.3906 & S_1 = 0.9943667 & A = 140. \\ \alpha_2 = 658.6132 & S_2 = 0.9048949 & B = 10. \end{array}$$

All the examples considered in this section refer to complete equal angular sets of mean values with the same statistical properties of terrestrial gravity anomalies (as far as such properties are known). The analysis of actual mean gravity anomalies is shown in the last paragraph. Besides being important in geodetic studies, gravity anomalies constitute a type of geophysical data with reasonably well known statistical properties, and their study here is meant to give the reader some idea of how effective are the ideas presented earlier when it comes to handling "real data" (or something resembling it).

3.2 Agreement between the Actual and the Theoretical Measures of the Sampling Errors

Table (3.1) lists side by side the theoretical percentage rms per degree of the sampling error according to (3.2) and the actual value of this percentage for two different sets of coefficients (i.e., from random sequences with different seeds). The coefficients were recovered using the quadratures formula

$$C_{n,n}^{\alpha} = \frac{1}{4\pi\beta_n} \sum_{i=0}^{n-1} \sum_{j=0}^{2n-1} \Delta g_{1,j} \int_{\sigma_{1,j}} \bar{Y}_{n,n}^{\alpha}(\theta, \lambda) d\sigma$$

This type of formulas has been discussed in paragraph (2.6). The simulated data consisted in full sets of $5^\circ \times 5^\circ$ mean anomalies obtained from harmonic coefficients complete to degree and order $N_{n,n} = 140$. The power above degree 140 in $5^\circ \times 5^\circ$ anomalies is negligible, according to the empirical power spectrum model that was used. The results shown here are fairly typical of similar tests conducted with other quadrature formulas, so the conclusions that can be drawn are likely to be valid for the analysis of area means by numerical quadratures in general. There is clear agreement between the theoretical and the actual rms of the errors, and not just the average rms of actual errors, but the actual rms of each trial as well. The agreement is close, and the reader will probably agree that to use a theoretical error measure that can predict the actual error so well is a meaningful way of quantifying the error.

3.3 Accuracies of Various Quadratures Formulas

Five quadratures formulas for area means have been studied: the first four of the type

$$\hat{C}_{n,n}^{\alpha} = \mu_n \sum_{i=0}^{n-1} \sum_{j=0}^{2n-1} \int_{\sigma_{1,j}} \bar{Y}_{n,n}^{\alpha}(\theta, \lambda) d\sigma \Delta g_{1,j}$$

Table 3.1

Comparison of actual versus theoretical percentage rms error per degree.

5° x 5° mean anomalies, $N_{max} = 140$, 0. mgal rms noise.

n	Actual, No 1 (seed=53218)	Actual, No 2 (seed=31765)	Average 1 and 2	Theoretical (expres- sion (3.2) x 100)
2	0.50	0.62	0.56	0.51
5	1.06	1.10	1.08	1.23
10	4.99	5.81	5.40	4.89
15	13.06	11.03	12.05	12.75
20	23.48	25.97	24.72	24.77
25	23.04	30.82	26.93	30.03
30	45.84	46.26	46.05	43.28
* 36 (N)	55.33	61.10	58.22	60.35
40	70.61	75.82	73.22	75.14
45	85.91	86.41	86.16	87.03
50	101.51	101.43	101.47	101.47

differ only in μ_n :

(a) $\mu_n = \mu_n^*$ the optimal de-smoothing factor given by expression (2.36) in paragraph (2.7):

$$(b) \mu_n = \frac{1}{4\pi\beta_n^2}$$

$$(c) \mu_n = \frac{1}{4\pi\beta_n}$$

$$(d) \mu_n = \frac{1}{4\pi}$$

(e) The optimal quadratures - type formula, in the least squares collocation sense (i.e., minimum combined error measure) for the given grid; signal and noise (paragraph (2.8)). The grid was equal angular in all five cases. Table (3.2) compares the percentage rms of the errors per degree for a 30° x 30° grid; table (3.3) corresponds to a 10° x 10° grid; and table (3.4) to a 5° x 5° grid. N_{max} was 100 for the first two tables, and 140 for the last. All these values are theoretical, computed in accordance to the formulas in paragraph (3.1). In all three cases noise is not present, thus errors are purely sampling errors.

Table (3.5) corresponds to a 5° x 5° grid and an uniform noise of 0. mgal. The effect of the noise has been included in the results.

3.3. Correlation coefficients for the various methods,
The correlation coefficient for the nth degree is defined as

where μ_n is

and

$$\rho_n = \left[\sum_{n=0}^N \sum_{\alpha=0}^1 \bar{C}_{nn}^{\alpha} \hat{C}_{nn}^{\alpha} \left(\sum_{n=0}^N \sum_{\alpha=0}^1 \bar{C}_{nn}^{\alpha} \right)^{-1} \left(\sum_{n=0}^N \sum_{\alpha=0}^1 \hat{C}_{nn}^{\alpha} \right)^{-1} \right]^{\frac{1}{2}} \quad (3.7)$$

and it is also equal to

$$\rho_n = \left[\frac{\int_{\sigma} \Delta g_n \Delta \hat{g}_n d\sigma}{\int_{\sigma} \Delta g_n^2 d\sigma \int_{\sigma} \Delta \hat{g}_n^2 d\sigma} \right]^{\frac{1}{2}} \quad (3.8)$$

This coefficient can be regarded either as a measure of the agreement between the actual and the recovered coefficients of the n th harmonic, or of between the n th harmonic Δg_n in the signal and the harmonic $\Delta \hat{g}_n$ that can be computed from the recovered coefficients. If the \bar{C}_{nn}^{α} could be seen as random variables with gaussian distribution, the interpretation of ρ_n would move along well-worn paths; however, as it was mentioned in paragraph (2.5), there are some unexpected problems when extending the idea of a gaussian random process to the sphere, so is better to choose another approach. One could regard the coefficients of the n th harmonic as the coordinates of a vector in $(2n+1)$ -dimensional euclidean space. The actual coefficients will define thus one vector, and the recovered coefficients another. Expression (3.7) then merely defines ρ_n as the scalar product of this two vectors. Likewise, expression (3.8) is that of the scalar product of two elements of a function space. The angle formed by these vectors is 0° when correlation is (maximum), and 90° for 0 correlation; a minimum correlation of -1 corresponds to the case when the vectors are equal but of opposite sense. The scalar product is independent of any scale factors that may multiply the vectors: it depends only on their mutual orientation. For this reason, the correlation coefficients are the same for the four quadratures formulas of the type

$$\bar{C}_{nn}^{\alpha} = \mu_n \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \Delta \bar{g}_{ij} \int_{\sigma} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma$$

because the difference between the errors for the same harmonic, predicted with two different formulas of this kind, consists in a scale factor $\mu_n^{(1)}/\mu_n^{(2)}$. Clearly, as the rms of the error increases with n , ρ_n decreases from almost 1 where the error is smallest (very low degrees), to below 0.5 where the error exceeds 90% (highest degrees analyzed).

Observing the percentual rms of the errors in the first four tables, it is easy to see that they by no means reach 100% as soon as the Nyquist frequency is reached ($n = N$), but that they remain substantially smaller than 100% even at degrees considerably higher than N ; this is in line with the conclusions in paragraph (1.3). The optimal estimator itself cannot have an error larger than 100%, be it due to sampling, noise, or both. Otherwise, a null estimator (one that predicts only zeroes) would be better than the optimal, which is not possible.

Table (3.7), compares the theoretical errors with zero noise (i.e., the sampling errors) of the collocation estimator obtained, first according to

σ_n^2 implied by R. Rapp's model (used in all the other tables), and then according to the "2L" model of C. Jelesky, both described in paragraph (3.1). This table is included here to give the reader an idea of how sensitive the theoretical error variances are to the empirical degree variances used to compute them. The "2L" model has considerably more power than Rapp's at high degrees, and this may be reflected in the somewhat larger errors in the corresponding column of the table.

Table 3.2					
Theoretical percentage rms error per degree. 30° x 30° mean anomalies, $N_{max} = 100$, 0. mgal rms noise.					
n	Optimal Estimator	μ_n^*	$\frac{1}{4\pi\beta_n}$	$\frac{1}{4\pi}$	$\frac{1}{4\pi\beta_n^2}$
2	12.51	13.17	13.95	15.86	13.57
4	25.09	25.60	26.91	36.12	28.72
6 (N)	58.77	65.13	68.03	67.18	101.41
10	92.82	96.01	194.08	97.13	882.98
12	98.44	99.27	915.49	105.44	17835.58

Table 3.3					
Theoretical percentage rms error per degree. 10° x 10° mean anomalies, $N_{max} = 100$, 0. mgal rms noise.					
n	Optimal Estim.	μ_n^*	$\frac{1}{4\pi\beta_n}$	$\frac{1}{4\pi}$	$\frac{1}{4\pi\beta_n^2}$
2	1.52	1.63	1.71	2.04	1.65
4	2.24	2.46	3.00	4.76	2.56
6	4.27	4.68	5.83	9.51	4.94
8	8.74	9.60	10.99	16.39	10.16
10	13.19	14.02	15.89	23.68	15.19
12	31.60	33.03	33.10	37.84	37.43
14	40.43	41.41	41.41	46.45	49.91
16	41.41	42.18	42.36	51.41	53.12
18 (N)	56.81	59.94	61.26	63.41	88.17
20	76.50	78.99	94.69	79.07	158.26
22	76.02	78.54	89.75	79.05	167.49
24	84.96	87.40	116.57	87.59	258.09
26	90.40	94.03	164.50	97.85	443.76
28	92.21	95.98	193.17	100.35	646.88
30	93.31	97.30	232.24	101.88	1010.13
32	95.80	98.30	303.89	103.49	1828.98
34	94.95	97.40	281.73	97.77	2619.98
36	96.87	98.08	485.72	98.45	8694.11

Table 3.4					
Theoretical percentage rms error per degree 5° x 5° mean anomalies, $N_{max} = 140$, 0. mgal rms noise.					
n	Optimal Estim.	μ_n^*	$\frac{1}{4\pi\beta_n}$	$\frac{1}{4\pi}$	$\frac{1}{4\pi\beta_n^2}$
2	0.47	0.49	0.51	0.59	0.50
6	1.36	1.44	1.71	2.61	1.48
12	9.74	10.11	10.40	12.27	10.37
18	15.14	15.55	16.35	21.20	16.46
24	30.58	31.06	31.21	36.30	34.89
30	42.88	43.28	43.28	49.04	53.32
36 (N)	57.62	59.15	60.34	62.67	85.90
42	72.16	73.51	80.25	74.72	137.19
48	82.15	83.62	101.19	83.82	216.98

Table 3.5					
Theoretical percentage rms per degree 5° x 5° mean anomalies, $N_{max} = 140$, 5. mgal rms noise.					
n	Optimal Estim.	μ_n^*	$\frac{1}{4\pi\beta_n}$	$\frac{1}{4\pi}$	$\frac{1}{4\pi\beta_n^2}$
2	8.74	8.81	8.84	8.83	8.85
6	9.33	9.34	9.34	9.43	9.41
12	32.80	32.88	33.79	33.02	35.28
18	35.94	36.04	36.54	36.25	39.86
24	53.32	53.47	56.73	53.49	66.99
30	61.97	62.14	67.15	62.15	87.22
36 (N)	71.98	72.62	82.81	72.63	122.67
42	81.55	82.17	102.91	82.54	181.19
48	88.00	88.72	124.09	89.54	271.87

Table 3.6		
Correlation factor per degree. 5° x 5° mean anomalies, $N_{max} = 140$, 0. mgal noise.		
n	Optimal Estim.	Simple quadratures
2	1.00	1.00
6	1.00	1.00
12	1.00	1.00
18	0.99	0.99
24	0.96	0.96
30	0.90	0.89
36 (N)	0.81	0.81
42	0.72	0.71
48	0.59	0.57

Table 3.7 Comparison between the theoretical errors (optimal estimator) when different covariance functions are assumed. $N_{max} = 140$, $5^\circ \times 5^\circ$ mean anomalies, 0. mgal noise.		
n	R. Rapp's variances	"2L" model
2	0.47	0.64
6	1.36	1.79
12	9.74	7.84
18	15.14	18.10
24	30.58	31.41
30	42.88	46.28
36 (N)	57.62	59.84
42	72.16	72.19
48	82.15	81.80

The results in the tables show that the optimal estimator errors are the smallest, as expected, and that the quadratures formula with the optimal de-smoothing factors is the best of the four simple quadratures formulas compared here, though it is not quite as good as the optimal estimator, also as expected. In the case of zero noise each of the three non-optimal quadratures formulas has errors that, in some region of the spectrum, are smaller than those for the other two; concretely, the de-smoothing factor $\mu_n = \frac{1}{4\pi\beta_n^2}$ works better for low degree harmonics (or $n \leq \frac{1}{3}N$), i.e., the percentage rms of the error is less than for the other two formulas there, the factor $\mu_n = \frac{1}{4\pi\beta_n}$ is best for middle harmonics (approximately $1/3N < n \leq N$), and $\mu_n = \frac{1}{4\pi}$ is best above the Nyquist frequency N . It is possible, therefore, to obtain a simple "composite" quadratures formula that combines the good properties of all three formulas, by defining its de-smoothing factor as follows:

$$\mu_n = \frac{1}{4\pi\eta_n} \text{ where } \eta_n = \begin{cases} \beta_n^2 & \text{if } 0 \leq n \leq 1/3N \\ \beta_n & \text{if } 1/3N < n \leq N \\ 1 & \text{if } n > N \end{cases} \quad (3.9)$$

This composite formula has been implemented in the version of subroutine HARMIN listed in Appendix B, through the subroutine can be easily changed to compute other quadrature formulas.

It is clear from the tables that, while better than all the others, the optimal formula is only marginally so: from a practical point of view, the simple "composite" formula (3.9) above is virtually as good, but it is much easier to implement and compute. Therefore, when analysing data of the type considered here (resembling mean anomalies with uniform noise on an equal angular grid), the quadratures formulas discussed the composite in particular, are about as good as any linear technique for estimating the coefficients, and also very easy to program and very efficient.

There may be cases, however, when this is not true. If the data set were very noisy and/or incomplete, or if the coefficients to be recovered were not those of the signal, but those of some complicated transformation of it (as in the case of satellite-to-satellite tracking data, where the signal depends on a combination of differences of radial and horizontal derivatives of the gravitational potential) so no simple integral formula like (1.4) exists that can be readily discretized into summations like (1.5) or (1.6), then simple quadrature formulas like those in section 1 would be of no real use, and the optimal estimator could provide the only practical way of obtaining the coefficients. Both collocation and least squares adjustment are very similar, as shown in section 2, and both can be implemented with reasonable efficiency if nothing simpler is available. More important, the optimal formulas provide a theoretical background on which one can build a coherent and comprehensive understanding of the other linear techniques for spherical harmonic analysis. It is, after all, because of the theory developed in the previous section that it has been possible to obtain the results shown in the preceding tables, results that constitute the factual basis for these considerations.

The techniques for setting up and inverting the variance-covariance matrix of the data are of interest in a number of estimation and filtering problems, besides harmonic analysis. The author hopes that the examples provided in this section will encourage the wider use of least squares collocation for the processing of large, global sets of data, both point values and area means. Creating the simulated mean anomalies with subroutine SSYNTH, obtaining the "weights" x_i^{nn} of the optimal quadratures formula with NORMAL (Appendix B), and recovering the \hat{C}_{nn}^α for comparison with the original \bar{C}_{nn}^α up to degree and order 72 (same 5000 coefficients) on a $5^\circ \times 5^\circ$ degree grid (about 2600 "data" values) took less than 20 seconds, central processor unit time, in the AMDHAL computer at Ohio State. To recover the harmonic coefficients up to degree and order 180 from a complete equal angular set of $1^\circ \times 1^\circ$ mean values would require less than two hours, using the same machine. Most of this time would be dedicated to creating and inverting $(C_{zz} + D)$. But in the optimal estimation and filtering of geoidal undulations, deflections of the vertical, and any other function of the gravity field estimable from (say) gravity anomalies, the fact that $(C_{zz} + D)$ can be set up and inverted efficiently transcends harmonic analysis. A major implication is that such extremely large global adjustments require a computational effort that is already within the reach of most researchers.

As mentioned already in section 2, matrix $(C_{zz} + D)$ may become poorer conditioned, i.e., its numerical inversion less stable, as the data distribution becomes denser. This tendency towards instability was noticed: when there was no noise ($D = 0$) the $R(m)$ matrices had to be regularized by adding a small positive constant k to each diagonal term (Colombo, 1979, par. (4.5)) before they could be inverted successfully. This constant, which in most cases was much smaller than the diagonal elements it was being added to, was 10^{-6} for $30^\circ \times 30^\circ$ and $10^\circ \times 10^\circ$, but had to be increased to 10^{-9} in the case of a $5^\circ \times 5^\circ$ grid. When noise was present, the nonzero diagonal elements in D

were sufficient to provide stability, and no regularization was needed. Because of this tendency to instability, the "band-limited" approach of paragraphs (2.13) to (2.17) may be preferable, whenever it can be properly applied.

As shown in Table (3.1), the theoretical and the actual sampling errors are almost the same in most cases. The propagated noise measure is very easy to compute in the case of uncorrelated noise, and can be added to the actual sampling error (variance) to get an estimate of the total error, both actual and theoretical. This estimate, where the sampling part is the result of a Montecarlo-like approach, is much easier to obtain than the theoretical one that involves setting up matrix $(C_{zz} + D)$, or at least the $R(m)$ matrices. In the case of equal angular data sets like those considered here, this empirical estimate is likely to be just as accurate, when it comes to judge the performance of any given type of harmonic analysis. Such estimate has been used to evaluate the likely errors in the potential coefficients obtained from $1^\circ \times 1^\circ$ mean anomalies using the quadratures formula $\bar{C}_{na}^\alpha = \frac{1}{4\pi\beta_n} \sum_{l=0}^{N-1} \sum_{m=0}^{n-1} \int_{\sigma_{1j}} \bar{Y}_n^\alpha(\theta, \lambda) d\sigma \Delta g_{1j}$. The Montecarlo method described in paragraph (3.1) was implemented with the help of subroutines SSYNTH and HARMIN, the error variance being equated to its average over three "trials" (three sets of coefficients created from different random sequences), C. Jekeli, also at O.S.U., who undertook this work as part of his own research, fitted a quartic to the percentage rms per degree thus obtained. When this was expressed as a function of the "normalized degree" n/N , the quartic fitted equally well the theoretical results for $30^\circ \times 30^\circ$, $10^\circ \times 10^\circ$, and $5^\circ \times 5^\circ$ presented in this section. Jekeli's quartic expression for the truncation error is (private communication):

$$\frac{\sigma_{\epsilon_n}}{\sigma_n} \times 100 = [(-16.19570 (\frac{n}{N}) + 30.34506) (\frac{n}{N}) + 40.29588] (\frac{n}{N})^2 \quad (3.10)$$

It is quite remarkable that such a complex phenomenon can be described satisfactorily by such a simple law.

The expansion of the simulated $1^\circ \times 1^\circ$ mean anomalies was complete up to degree and order $N_{max} = 300$. Creating (or analysing) area mean values up to degree and order $N_{max} = 300$ required about 50 seconds o.p.u. time using double precision arithmetic in the AMDHAL computer at O.S.U.

Table 3.8 shows the actual percentage rms sampling error per degree as computed in one of the trials, and the percentage rms propagated noise (theoretical) corresponding to a 1 mgal rms noise in the data. Clearly, the errors are much smaller than for any of the cases considered previously: this improvement is due to the finer sampling (the sampling error tends to zero as the area of the blocks tends to zero). The data, however, tends to be noisier when averaged on smaller blocks, so the propagated noise may increase. For a given rms error in the data, multiply the number in the "propagated noise" column by this rms (in mgals) to obtain the corresponding percentage. These numbers are only valid for the estimator where $\mu_n = \frac{1}{4\pi\beta_n}$. Repeated trials with different random coefficients resulted in much the same percentages for the sampling errors, so these values are probably fairly typical.

Table 3.8 Percentage rms errors per degree for a $1^\circ \times 1^\circ$ equal angular grid. $\mu_a = \frac{1}{4\pi\beta_a}$, $N_{max} = 300$, $\epsilon \Delta g = 1$ mgal		
n	actual sampling	propagated error (theor.)
2	0.01	0.47
10	0.10	0.72
20	0.46	1.39
30	0.98	1.72
40	1.55	1.93
50	2.68	2.07
60	4.71	2.20
70	6.94	2.34
80	9.43	2.47
90	13.51	2.62
100	14.79	2.77
110	17.98	2.96
120	25.06	3.08
130	25.33	3.26
140	30.87	3.43
150	37.02	3.62
160	44.17	3.81
170	44.43	4.03
180 (N)	53.07	4.24
190	61.87	4.47
200	65.90	4.73

A possible way of bringing the sampling error down is to use weighted area means of the form

$$\Delta \bar{g}_{i,j} = \sum_{k=1}^{N_{i,j}} \Delta g_{i,j}^{(k)} W_k$$

where $\Delta g_{i,j}^{(k)}$ is the k th measurement inside the block $\sigma_{i,j}$, and where the W_k are functions of the distance of $\Delta g_{i,j}^{(k)}$ to the center of the block. If the W_k decay gently towards the border of the area element, the resulting weighted means will be smoother than the ordinary area means considered so far (all $W_k = N_{i,j}^{-1}$) and their harmonic content above the Nyquist frequency will be attenuated. Consequently, the harmonics below N can be recovered with less sampling error. This idea certainly deserves further study. Obviously, it is applicable only in those cases where the original measurements $\Delta g_{i,j}^{(k)}$ are available.

3.4 The Analysis of a Global Data Set of 5° x 5° Mean Anomalies

As a final demonstration of the use of optimal estimators, this paragraph presents some results obtained by analysing a real data set, consisting of 5° x 5° mean gravity anomalies. These anomalies were obtained from a global set of 1° x 1° mean values created by R. Rapp and associates (Rapp, 1978) from the combination of land and gravity measurements, satellite altimetry over the oceans, and the GEM-9 satellite model (Lerch et al., 1979).

The 5° x 5° values, $\Delta g_{(5^\circ)}$, were obtained from the 1° x 1° values, $\Delta g_{(1^\circ)}$, using the formula

$$\Delta g_{(5^\circ)} = \frac{1}{25} \sum_{i,j} \Delta g_{(1^\circ)}_{i,j}$$

where $\sum_{i,j}$ represents summation over all 1° x 1° blocks inside a 5° x 5° block. The variances of the $\Delta g_{(5^\circ)}$ were obtained from the formula

$$\sigma^2 \Delta g_{(5^\circ)} = \frac{1}{(25)^2} \sum_{i,j} \sigma^2 \Delta g_{(1^\circ)}_{i,j}$$

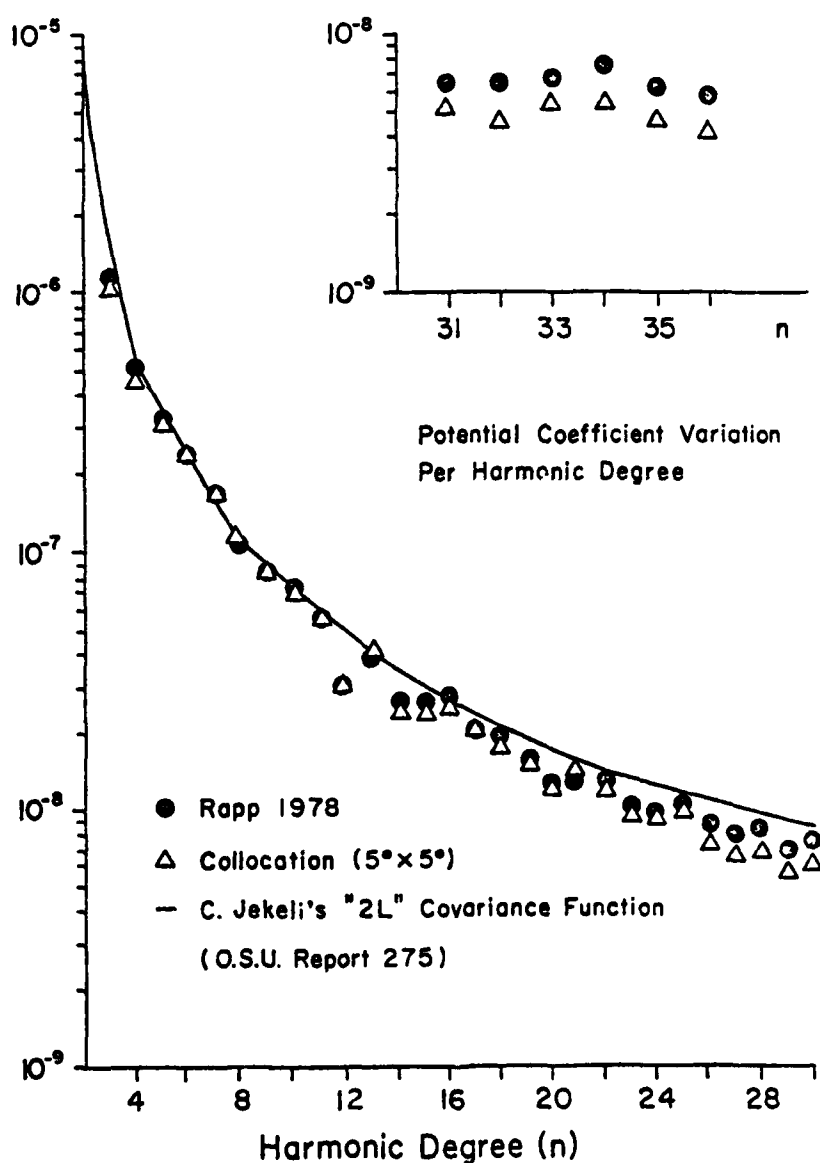
where the assumption has been made that the errors in the 1° x 1° values were uncorrelated, which is not quite true, as the values used were the product of an adjustment. The variances $\sigma^2 \Delta g_{(5^\circ)}$ were different from 5° block to 5° block, but they were "homogenized" as described in paragraph (2.9), in order to obtain a (C₂₂ + D) matrix that was easily invertible and resulted in a quadratures-type estimator inexpensive to implement.

Figure (3.1) shows a comparison between the collocation solution, complete to degree and order 36, and the same coefficients obtained by R. Rapp by numerical quadratures from the original 1° x 1° values. The figure shows the rms per degree ($\sqrt{\sigma_n^2/(2n+1)}$) for potential coefficients ($\sigma_n^2(T) = \sigma_n^2(\Delta g) \gamma^{-2} (n-1)^{-2}$, $\gamma = 979800$ mgal). The circles correspond to Rapp's results, and the triangles to those obtained using collocation as already explained. Because the grid used by Rapp is much finer, the corresponding results are likely to be less affected by sampling errors, at least up to degree 36, than those obtained from the 5° x 5° anomalies; for this reason Rapp's rms values are regarded here as the "true" ones. The solid line corresponds to C. Jeleski's "2L" model for the σ_n^2 , used here to obtain the optimal estimator. It is interesting to notice that the triangles follow the circles (or "true" values) rather than the line. A common concern among those using this type of estimators is to what extent the "a priori" power spectrum or covariance function used to set up the estimator may "bias" the results by forcing the spectrum of the output to resemble the "a priori" spectrum. Here there is little evidence of such a "bias".

In addition to Rapp's coefficients, those of GEM-10B (Lerch, 1980) were also used for comparison. The "collocation" values follow them very closely too (the three sets of results agree, in fact, very well with each other). The

GEM values are not shown in the figure, to have a cleaner picture. In any event, the "collocation" values compare very well with the other two, and in the higher degrees (31 to 36) where the divergence between Rapp's model and collocation is largest, GEM-10B fits right in between them. The results shown here were presented in a previous paper by the author (Colombo, 1979-b).

Figure 3.1



4. Covariances Between Area Means

In what follows, the expression "point" covariance refers to

$$\text{cov}(u(P), v(P')) = \sum_{n=0}^{\infty} c_n^{u,v} P_n(\psi_{PP'}) = M\{u(P), v(P')\} \quad (4.1)$$

The covariance between the area means of two functions u and v is related to the "point" covariance function by the following integral relationship (Sjöberg, 1978):

$$\text{cov}(\bar{u}_{ij}, \bar{v}_{kl}) = \frac{1}{\Delta_{ij}\Delta_{kl}} \int_{\sigma_{ij}} d\sigma \int_{\sigma_{kl}} \text{cov}(u(P), v(P')) d\sigma' \quad (4.2)$$

where P belongs to the area block σ_{ij} , and P' to σ_{kl} , while

$$\bar{u}_{ij} = \frac{1}{\Delta_{ij}} \int_{\sigma_{ij}} u(P) d\sigma \quad (4.3)$$

is the average of u over the block σ_{ij} of area

$$\Delta_{ij} = \Delta\lambda (\cos(\theta_i) - \cos(\theta_i + \Delta\theta_i)) \quad (4.4)$$

$\Delta\theta_i, \Delta\lambda$ are the colatitude and longitude spans of the blocks in the "row" between colatitudes θ_i and $\theta_i + \Delta\theta_i$. To simplify the discussion, $\Delta\theta_i = \Delta\theta$ is assumed constant; extension to the more general case where $\Delta\theta_i$ varies from row to row is straightforward.

4.1. Derivation of an Approximate Formula for the Covariance¹

Expression (4.2) can be computed by numerical quadratures (Rapp, 1977). Because of the double integral, this is a very laborious process, and it is not practical if many covariance values have to be found, as in the case of large data sets. A more efficient alternative is needed.

Replacing the covariance function in the integrand of (4.2) with its Legendre expansion

$$\text{cov}(\bar{u}_{ij}, \bar{v}_{kl}) = \frac{1}{\Delta_{ij}\Delta_{kl}} \int_{\sigma_{ij}} d\sigma \int_{\sigma_{kl}} \sum_{n=0}^{\infty} c_n^{u,v} P_n(\psi_{PP'}) \quad (4.5)$$

Here $\psi_{PP'} = \cos^{-1}(\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\lambda - \lambda'))$ is the spherical distance between points $P \equiv (\theta, \lambda)$, and $P' \equiv (\theta', \lambda')$ in the unit sphere, while

$$c_n^{u,v} = \sum_{m=0}^n \bar{C}_{nm}^u \bar{C}_{nm}^v + \bar{S}_{nm}^u \bar{S}_{nm}^v$$

is the n th degree variance of the cross-spectrum of u and v . As shown in the Appendix, the order of summation and integration can be reversed.

¹ In this section, expression $P_n(\psi_{PP'})$ is shorthand notation for $P_n(\cos\psi_{PP'})$.

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = \frac{1}{\Delta_{1j}\Delta_{k1}} \sum_{n=0}^{\infty} \frac{c_n^{u,v}}{2n+1} \int_{\sigma_{1j}} d\sigma \int_{\sigma_{k1}} P_n(\psi_{pp'}) d\sigma' \quad (4.6)$$

Let $P_{nm}(\cos \theta)$ be the fully normalized associate Legendre function of the first kind of degree n and order m . Applying the Summation Theorem for such functions, the $P_n(\psi_{pp'})$ can be replaced in the integrand as shown below.

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = \frac{1}{\Delta_{1j}\Delta_{k1}} \sum_{n=0}^{\infty} \frac{c_n^{u,v}}{2n+1} \int_{\sigma_{1j}} d\sigma \int_{\sigma_{k1}} \sum_{m=0}^n \bar{P}_{nm}(\cos \theta) \bar{P}_{nm}(\cos \theta') \cos m(\lambda - \lambda') d\sigma' \quad (4.7)$$

The sum in the general term of the series above has a finite number of terms $(n+1)$, so it is valid to exchange summation and integration once more:

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = \frac{1}{\Delta_{1j}\Delta_{k1}} \sum_{n=0}^{\infty} \frac{c_n^{u,v}}{2n+1} \sum_{m=0}^n \int_{\sigma_{1j}} d\sigma \int_{\sigma_{k1}} \bar{P}_{nm}(\cos \theta) \bar{P}_{nm}(\cos \theta') \cos m(\lambda - \lambda') d\sigma' \quad (4.8)$$

Writing out the area integrals as colatitude and longitude integrals:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = & \frac{1}{\Delta_{1j}\Delta_{k1}} \sum_{n=0}^{\infty} \frac{c_n^{u,v}}{2n+1} \sum_{m=0}^n \int_{\theta_1}^{\theta_1+\Delta\theta_1} \bar{P}_{nm}(\cos \theta) \sin \theta d\theta \cdot \\ & \cdot \int_{\theta_k}^{\theta_k+\Delta\theta_k} \bar{P}_{nm}(\cos \theta') \sin \theta' d\theta' \int_{\lambda_j}^{\lambda_j+\Delta\lambda} d\lambda \int_{\lambda_1}^{\lambda_1+\Delta\lambda} \cos m(\lambda - \lambda') d\lambda' \end{aligned} \quad (4.9)$$

Calling $\Delta_1 \equiv \Delta_{1j}$, as all blocks in the same "row" have the same area, we introduce two functions,

$$F(m) = \begin{cases} \Delta\lambda^2 & \text{if } m=0 \\ (2/m^2)(1-\cos m\Delta\lambda) & \text{otherwise,} \end{cases} \quad (4.10-a)$$

and

$$I_{nm,i} = \left(\frac{c_n^{u,v}}{2n+1} \right)^{\frac{1}{2}} \frac{1}{\Delta_1} \int_{\theta_1}^{\theta_1+\Delta\theta_1} \bar{P}_{nm}(\cos \theta) \sin \theta d\theta \quad (4.10-b)$$

Then expression (4.9) can be written, after integrating and reordering terms, as follows:

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = \sum_{n=0}^{\infty} F(m) \sum_{m=-n}^n I_{nm,i} I_{nm,k} \cos m(\lambda_i - \lambda_j) \quad (4.11-a)$$

This regrouping of the series is valid because, as shown in the Appendix, the series is absolutely convergent. The integrals in the definition of $I_{nm,i}$ (expression (4.10-b)) can be calculated very accurately and efficiently with recursive formulae obtained by M. K. Paul (1979). Regarding j as being fixed, the last expression is also that of the Fourier series of cosines of $\text{cov}(u_{1j}, v_{k1})$, with amplitudes

$$a_n^{i,k} = F(m) \sum_{m=-n}^n I_{nm,i} I_{nm,k} \quad (4.11-b)$$

and phases $\phi_n = -m\lambda_j$.

While (4.11-a, b) are valid for all values of λ_j in the interval $0 \leq \lambda_j \leq 2\pi$, this expression has to be calculated only for $\lambda_j = 0, \Delta\lambda, 2\Delta\lambda, \dots, j\Delta\lambda$. According to the sampling theorem for ordinary Fourier Series, at such regularly spaced λ_j expression (4.11-a) takes precisely the same values as the finite sum

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = \sum_{n=0}^N \hat{a}_n^{1k} \cos m(\lambda_j - \lambda_1) \quad (4.12)$$

where $N = \pi/\Delta\lambda$

$$\hat{a}_n^{1k} = \sum_{h=0}^{\infty} a_{2Nh+n}^{1k} + \sum_{h=1}^{\infty} a_{2Nh-n}^{1k} \quad (4.13)$$

To calculate the covariances, (4.11-a, b) and (4.13) have to be truncated, excluding all harmonics above some degree N_{\max} ; (4.12) becomes:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) \approx \text{cov}(\bar{u}_{1j}, \bar{v}_{k1})_{N_{\max}} &= \sum_{n=0}^N \left[\sum_{h=0}^K \sum_{n=n}^{N_{\max}} (I_{n, 2Nh+n, 1})(I_{n, 2Nh+n, k}) \right. \\ &\quad \left. + \sum_{h=1}^K \sum_{n=n}^{N_{\max}} (I_{n, 2Nh-n, 1})(I_{n, 2Nh-n, k}) \right] F(m) \cos m(\lambda_j - \lambda_1) \end{aligned} \quad (4.14)$$

where $(2K+1)N \leq N_{\max}$.

4.2. Choosing N_{\max}

To calculate the values of covariances between equal angular mean anomalies, $\bar{\Delta g}$, the value of N_{\max} could be chosen so that the percentage error

$$v = \left| \frac{\text{cov}(\bar{\Delta g}_{1j}, \bar{\Delta g}_{k1})_Q - \text{cov}(\bar{\Delta g}_{1j}, \bar{\Delta g}_{k1})_{N_{\max}}}{\text{cov}(\bar{\Delta g}_{1j}, \bar{\Delta g}_{k1})_Q} \right| \times 100$$

($\text{cov}(\bar{\Delta g}_{1j}, \bar{\Delta g}_{k1})_Q$ being computed by numerical quadratures) does not exceed a prescribed upper bound ϵ . The smallest N_{\max} that meets this condition increases towards the poles, because the decrease in area of equal angular blocks with latitude means that the averages have a high frequency content that increases accordingly. On the other hand, the absolute values of the integrals of the \bar{P}_{na} , and therefore the Fourier coefficients a_n in (4.12), decrease quite fast with increasing order m near the poles, so their contribution soon becomes insignificant. This is fortunate, because the need for lengthier calculations for each Fourier coefficient nearer to the poles can be offset by the existence of fewer coefficients there. In fact, because of the finite arithmetic of digital computers, all \hat{a}_n for blocks less than 30° from the poles are rounded off to zero for m considerably less than N in the cases presented here. Because of this, calculations near the poles can be less laborious than close to the equator, in spite of the larger N_{\max} . To take advantage of this in the programming of (4.14), the \hat{a}_n coefficients were not computed above the first m for which the following condition was met:

$$\frac{a_{n-1}^2}{\sum_{\ell=0}^{n-1} a_{\ell}^2} \leq \delta \quad \text{and} \quad \frac{a_n^2}{\sum_{\ell=0}^n a_{\ell}^2} \leq \delta,$$

where $\delta \leq 10^{-12}$. This ensured no change in the first six significant figures of the result when compared to the case where no coefficients were ignored, but there were great savings in computing.

4.3. Numerical Examples

To verify the accuracy of the truncated series in expression (4.14), mean gravity anomaly covariances were computed both with this formula and by numerical quadratures. All calculations were carried out in double precision (32 bytes), some of those for expression (4.14) being repeated in extended precision (64 bytes). The agreement between both sets of results was better than 6 significant figures, suggesting that both expression (4.14) and Paul's recursives (used to obtain the $I_{nn,1}$) are quite stable numerically, at all latitudes.

The covariances computed were between a mean anomaly in a fixed block and all the anomalies in the same row of blocks (i.e., all blocks bounded by the same parallels), calculations being done for several rows, at close intervals from equator to pole. Results for only a few of those will be shown here, because they are typical of the rest. Both a 5° and a 1° grid (equal angular) were studied. The results were compared to those obtained by numerical quadratures, and by the approximate formula

$$\text{cov}(\bar{\Delta g}_{1j}, \bar{\Delta g}_{k1}) \approx \sum_{n=0}^{N_{\max}} \beta_{n,1} \beta_{n,k} c_n^{\Delta \epsilon, \Delta \epsilon} P_n(\psi_{YY'}) \quad (4.15)$$

where Y and Y' are the center points of σ_{1j} and σ_{k1} , while

$$\beta_{n,1} = \frac{1}{1 - \cos \psi_0} \frac{1}{2n+1} [P_{n-1}(\cos \psi_0) - P_{n+1}(\cos \psi_0)]$$

is the n th degree Pellinen smoothing factor (the formula is Meissl's), and

$$\psi_0 = \cos^{-1} \left[\frac{\Delta \lambda}{2\pi} (\cos(\theta_1 + 1) - \cos \theta_1) + 1 \right]$$

This formula gives the covariances of averages on circular blocks of the same area as that of the equal angular blocks in the row between colatitudes θ_1 and θ_{1+1} . These covariances are used sometimes as approximations to the equal angular covariances between mean values.

The numerical quadratures technique consists in the following: (a) subdividing each block with a grid of k equally spaced latitude and k equally spaced longitude lines; (b) computing the covariances between point gravity anomalies at the nodes of each subdivision (there are k^4 different pairs of nodes to be considered); (c) obtaining the approximate value of the covariance between mean anomalies as

$$\text{cov}(\overline{\Delta g_{ij}}, \overline{\Delta g_{kl}}) = \sum_{n=0}^{k-1} \sum_{m=0}^{k-1} \sum_{r=0}^{k-1} \sum_{s=0}^{k-1} \text{cov}(\Delta g_{ij}^{mn}, \Delta g_{kl}^{rs}) \quad (4.16)$$

where $m, n, r,$ and s are indexes that identify the elements of each pair of nodes in the subdivisions of σ_{ij} and σ_{kl} . The covariances between point anomalies such as Δg_{ij}^{mn} and Δg_{kl}^{rs} were obtained using a two term model for the degree variances of the anomalies:

$$c_n^{\Delta g, \Delta g} = 18.3906 \frac{(n-1)}{(n+100)} 0.9943667^{n+2} + 658.6132 \frac{(n-1)}{(n+20)(n-2)} 0.9048949^{n+2}$$

This is the "2L" model of C. Jekeli (1978), and has the advantage that the value of the point covariance function can be computed using finite recursion formulas (Moritz, 1977). The same degree variances were used in (4.14) and (4.15). The number of point covariances being very large (k^4), a table with entries spaced at $\Delta\psi = 0.05^\circ$ intervals was created first, the required values being obtained by linear interpolation from this table. Numerical tests showed that $k = 10$ was large enough for both 5° and 1° blocks, because doubling this number resulted in a change of less than 0.2% in the values given by (4.16). Reducing the interval $\Delta\psi$ from 0.05° to 0.005° had a negligible effect also, therefore the values obtained with (4.16) are probably accurate enough to test those given by (4.14). The only exception was the "polar" row, where the equal angular blocks are, in fact, triangles with a common vertex at the pole. Both with 5° and 1° blocks the discrepancies between (4.14) and (4.16) were large (more than 30%), regardless of how large a k , how small a $\Delta\psi$, or how big a N_{\max} were chosen. The probable explanation is that the pole is given undue weight in (4.16) because it is treated as a whole row, instead of as a single point. For this reason, the numerical examples presented here stop at the row immediately below the pole.

The first two tables show the covariances between 5° mean anomalies in the row between latitudes 0° and 5° (just above the equator) and in the row between latitudes 80° and 85° (one below the pole). The error is at most 8%, though much less in most cases, and $N_{\max} = 180$ in each table. Under "Pellinen" one finds the values obtained using (4.15) with due regard for the change in block areas with latitude. While there is very good agreement near the equator, there is no resemblance at all close to the pole to the other values listed.

Table 4.1. Comparison between covariances of 5° mean anomalies $\overline{\Delta g}$ computed with expressions (4.14), (4.16) and (4.15), respectively. $N_{\max} = 180$. Row between 0° and 5°.

Expression (4.14)	Numerical	Pellinen	Block No.
250.53	253.77	251.78	1
148.80	149.46	151.03	2
193.93	93.95	93.90	3
57.16	57.15	57.10	4
31.80	31.76	31.67	5
13.94	13.93	13.86	6
-18.09	-18.09	-18.08	12
9.12	9.12	9.11	24
-13.68	-13.68	-13.65	36

Table 4.2. Comparison between covariances of 5° mean anomalies $\overline{\Delta g}$ computed with expressions (4.14), (4.16) and (4.15), respectively. $N_{\max} = 180$. Row between 80° and 85°.

Expression (4.14)	Numerical	Pellinen	Block No.
418.60	437.23	835.53	1
329.03	318.27	800.09	2
220.91	229.17	709.03	3
191.75	196.40	592.19	4
179.79	179.70		5
165.58	168.46		6
120.60	123.73		12
69.89	73.27		24
54.66	58.01	149.40	36

The next three tables show results for 1° blocks. Here the numerical method was conducted with the same $\Delta\psi$ and k as in the case of the 5° grid. Results for rows between latitudes 0° and 1°, 45° and 46°, and 88° and 89° are shown. In tables 4.3 and 4.4 the discrepancies between (4.14) and (4.16) stay below 1%; this increases to about 5% near the pole (table 4.5). N_{\max} is 300 for the equatorial row, and rises to 400 from 45° on. As with 5° blocks, the "Pellinen" values are quite close to these of (4.14) and (4.16) near the equator, but become very different near the pole.

Table 4.3. Comparison between covariances of 1° mean gravity anomalies computed with (4.14) and (4.16). $N_{\max} = 300$.
Row between 0° and 1° .

Expression (4.14)	Numerical	Block No.
849.68	855.16	1
411.32	410.30	2
220.35	219.63	3
181.84	181.50	4
163.33	163.21	5
149.11	149.15	6
1.07	1.08	31
-16.79	-16.79	61
1.99	1.99	91
8.54	8.53	121
-4.79	-4.79	151
-14.39	-14.38	181

Table 4.4. Comparison between covariances of 1° mean gravity anomalies computed with (4.14) and (4.16). $N_{\max} = 400$.
Row between 45° and 46° .

Expression (4.14)	Numerical	Block No.
952.25	959.17	1
531.15	531.97	2
275.55	274.65	3
210.81	210.46	4
185.36	185.57	5
170.82	171.08	6
27.79	27.79	31
-14.37	-14.37	61
-18.01	-17.01	91
-8.27	-8.28	121
-1.01	-2.92	151
1.39	1.39	181

Table 4.5. Comparison between covariances of 1° mean gravity anomalies computed with (4.14), (4.16), and (4.15), respectively. $N_{\max} = 400$. Row between 88° and 89° .

Expression (4.14)	Numerical	Pellinen	Block No.
1137.97	1187.94	1795.88	1
1135.79	1184.15	1795.02	2
1129.28	1177.99	1792.45	3
1118.55	1155.08	1788.18	4
1103.76	1131.34	1782.25	5
1085.15	1102.85	1774.68	6
408.59	443.12		31
248.27	257.88		61
207.19	210.64		91
189.65	192.55		121
181.93	184.53		151
179.65	179.86	368.60	181

Calculations were carried out in the Ohio State University's AMDHAL 470 V/6-II computer, using the FORTRAN H EXTENDED compiler, and double precision. The computing times for obtaining all 37 different covariances between the elements of a row of 5° mean anomalies, and all 181 covariances in a row of 1° mean anomalies, using (4.14) and (4.16) are listed in table 4.6 for comparison. The integrals of the Legendre functions required by (4.14), and the table of point anomalies covariances needed in (4.16), were precomputed and stored in core memory arrays, so the times given here do not include the determination of those auxiliary values. In most ordinary applications of these formulas, those quantities can be read from disk or tape whenever needed, because they are the same for a whole variety of problems. Clearly, using (4.14) can be orders of magnitude more efficient than using (4.16), while the accuracy is much the same. In fact, accuracy is probably better at the polar rows with (4.14) than with (4.16), because the latter seems to have problems handling triangular blocks. Finally, not all of the time-saving properties of (4.14) were exploited in the computer program used to calculate it, so there is scope for some improvement in efficiency beyond that shown in the table. Notice the time saved in the $88^\circ - 89^\circ$ row thanks to the neglect of terms in (4.14) that become too small near the poles, as explained in paragraph (4.2).

Table 4.6. Comparative efficiencies of algorithms based on expressions (4.14) and (4.16).

Block Size	Row	N _{max}	Expression (4.14)	Numerical Quadratures
5°	0° - 5°	180	0.1 sec.	23.3 sec.
5°	80° - 85°	180	0.1 sec.	23.3 sec.
1°	0° - 1°	300	0.35 sec.	113.3 sec.
1°	45° - 46°	400	0.43 sec.	113.3 sec.
1°	88° - 89°	400	0.06 sec.	113.3 sec.

4.4. Covariances Between Mean Values and Point Values

The prediction by least squares collocation of mean values from point values, or vice versa, requires finding the corresponding "mixed" covariances. In such case, formula (4.2) becomes

$$\text{cov}(\bar{u}_{ij}, v(\theta, \lambda)) = \frac{1}{\Delta_{ij}} \int_{\sigma_{ij}} \text{cov}(u(P'), v(P)) d\sigma \quad (4.17)$$

where $P' \equiv (\theta', \lambda') \in \sigma_{ij}$, $P \equiv (\theta, \lambda)$, $v(P)$ is the "point" value of v , and \bar{u}_{ij} is the average over the i, j block of the grid, as before. Following a similar reasoning, one arrives at a formula that corresponds to (4.14), except that only one area integral has to be considered. The new expression is:

$$\begin{aligned} \text{cov}(\bar{u}_{ij}, v(\theta, \lambda)) \sim & \frac{1}{\Delta_{ij}} \sum_{n=0}^N \left[\sum_{h=0}^K \sum_{n=n}^{N_{\max}} \left(\frac{c_n^{u,v}}{2n+1} \right)^{\frac{1}{2}} (I_{n,2Nh+n,1} \bar{P}_{n,2Nh+n}(\cos \theta)) \right. \\ & \left. + \sum_{h=1}^K \sum_{n=n}^{N_{\max}} \left(\frac{c_n^{u,v}}{2n+1} \right)^{\frac{1}{2}} (I_{n,2Nh-n,1} \bar{P}_{n,2Nh-n}(\cos \theta)) \right] \left[A(m) \cos m(\lambda_j - \lambda) \right. \\ & \left. + B(m) \sin m(\lambda_j - \lambda) \right] \quad (4.18) \end{aligned}$$

where

$$A(m) = \begin{cases} \Delta\lambda & \text{if } m = 0 \\ (\cos m\Delta\lambda - 1)/m & \text{otherwise} \end{cases} \quad B(m) = \begin{cases} 0 & \text{if } m = 0 \\ (\sin m\Delta\lambda)/m & \text{otherwise} \end{cases}$$

and the $I_{n,n,1}$, the $c_n^{u,v}$, and the N , K , and N_{\max} are as in (4.14). Expression (4.18) assumes that the area means belong to a grid with constant $\Delta\lambda$. If the point values are also on a grid, and if this grid is congruent with that of the mean anomalies, implementation of (4.18) is quite efficient. In fact, the speed of a good algorithm for doing this should be much the same as that of one for implementing (4.14). On the other hand, computing the same covariances by numerical quadratures is k^2 times faster in this "mixed" case than it was in the previous one, because there is only one area integration involved. Assuming $k = 10$, as in the previous examples, then expression (4.18) should be only 2-3 times faster than numerical quadratures for 5° mean values, and from 4 to 15 times faster in the case of 1° averages.

Both the normalized Legendre function and their integrals are needed for (4.18). They can be precomputed and stored on disk or tape until needed. In this study the following recursive relationships were used to generate their values:

$$\begin{aligned} \bar{P}_{n-1,n-1}(\cos \theta) = & \left[\frac{(2n-1)(2n-3)}{(n-m)(n+m-2)} \right]^{\frac{1}{2}} (\cos \theta \bar{P}_{n-2,n-1}(\cos \theta) \\ & - \left[\frac{(2n-1)(n+m-3)(n-m-1)}{(2n-5)(n+m-2)(n-m)} \right]^{\frac{1}{2}} \bar{P}_{n-3,n-1}(\cos \theta)), \quad m \neq n \quad (4.19a) \end{aligned}$$

$$\bar{P}_{n-1,n-1}(\cos \theta) = \left[\frac{(2n-1)}{(2n-2)} \right]^{\frac{1}{2}} \sin \theta \bar{P}_{n-2,n-2}(\cos \theta) \quad (4.19b)$$

$$\begin{aligned} \int_{\theta_1}^{\theta_2} \bar{P}_{n-1,n-1}(\cos \theta) \sin \theta d\theta = & -\frac{1}{n} \left[\frac{(2n-1)(2n-3)}{(n-m)(n+m-2)} \right]^{\frac{1}{2}} \sin^2 \theta \bar{P}_{n-2,n-1}(\cos \theta) \Big|_{\theta_1}^{\theta_2} + \\ & + \frac{(n-3)}{n} \left[\frac{(2n-1)(n+m-3)(n-m-1)}{(2n-5)(n+m-2)(n-m)} \right]^{\frac{1}{2}} \int_{\theta_1}^{\theta_2} \bar{P}_{n-3,n-1}(\cos \theta) \sin \theta d\theta \quad (4.20a) \end{aligned}$$

$$\begin{aligned} \int_{\theta_1}^{\theta_2} \bar{P}_{n-1,n-1}(\cos \theta) \sin \theta d\theta = & \frac{1}{2n} \left[\frac{(2n-1)}{(n-1)(n-2)} \right]^{\frac{1}{2}} \sin^2 \theta \bar{P}_{n-2,n-3}(\cos \theta) \Big|_{\theta_1}^{\theta_2} + \\ & + \frac{1}{2n} \left[\frac{(n-1)(2n-1)(2n-3)}{(n-2)} \right]^{\frac{1}{2}} \int_{\theta_1}^{\theta_2} \bar{P}_{n-3,n-3}(\cos \theta) \sin \theta d\theta \quad (4.20b) \end{aligned}$$

if θ is not very small, or

$$\begin{aligned} \int_{\theta_1}^{\theta_2} \bar{P}_{n-1,n-1}(\cos \theta) \sin \theta d\theta = & - \left[\frac{(2n-1)(2n-3)\dots 3}{(2n-2)(2n-4)\dots 4} \right]^{\frac{1}{2}} \sin^{n+1} \theta \cdot \\ & \cdot \left[\frac{1}{n+1} + \frac{1}{2} \frac{\sin^2 \theta}{n+3} + \frac{1 \cdot 3}{2 \cdot 4} \frac{\sin^4 \theta}{n+5} + \frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \frac{\sin^6 \theta}{n+7} + \dots \right] \Big|_{\theta_1}^{\theta_2} \quad (4.20c) \end{aligned}$$

if θ is very small. Here $y(\theta) \Big|_{\theta_1}^{\theta_2} \equiv y(\theta_2) - y(\theta_1)$. The recursive formulas for the integrals of the Legendre functions were derived by M.K. Paul (1978); the author has been fortunate enough to have available a FORTRAN subroutine programmed by Paul, and kindly sent by him to Professor R.H. Rapp of the Department of Geodetic Science at O.S.U. The results reported here have been made possible by, and bear witness to, the great numerical stability of Paul's formulas.

5. Conclusions

The relationships between spherical harmonic series and Fourier series, coupled to the symmetries of spherical grids, permit the development of efficient algorithms for numerical analysis of data regularly sampled on the sphere.

The algorithms presented in section 1 for implementing numerical quadratures are efficient enough to allow the analysis of 64800 $1^\circ \times 1^\circ$ mean values through degree 180 in less than 20 seconds, and the summation of the 90000 terms of a harmonic series complete through degree 300, on a full $1^\circ \times 1^\circ$ equal angular grid, in less than 1 minute. The analysis of large global data sets to a very high resolution is a relatively trivial operation with modern digital computers.

The principles and ideas behind the optimal estimators of section 2 provide a rational basis for the study of linear techniques for spherical harmonic analysis, both optimal and non-optimal. The error measure introduced in this section is shown to be a very reasonable way of evaluating the estimation error, as illustrated by the results listed in table (3.1).

The optimal estimators themselves are reasonably easy to compute and use, particularly when they are of the quadratures type, which happens under the fairly general conditions discussed in section 2. Even when such conditions are not present, as in the case of scattered data, the problem still has a structure strong enough to allow efficient algorithms for creating and inverting the normal matrix.

The separation of the problem of estimating the coefficients (by least squares collocation or by least squares adjustment) according to the order m of the coefficients, allows both for efficiency and for numerical stability. Even if the total number of unknown coefficients is very large, the largest matrix to be inverted is of dimension N , instead of $O(N^2)$, as it would be if the problem could not be separated in this way.

All the algorithms presented here, when the grid is complete and regular, are well suited to parallel processing.

In the case of full grids of mean values with even noise, the results of section 3 suggest that the optimal "collocation" estimator can be approximated very closely by a much simpler quadratures-type formula, the "composite formula" (3.9). The search for simple, near-optimal estimators is just as important, from a practical point of view, as the search for efficient algorithms for obtaining the optimal estimators themselves. This is a topic that certainly deserves further research.

The methods for creating and inverting the normal matrix, that make possible to find optimal estimators for large data sets, have application outside

spherical harmonic analysis, in all areas of estimation, filtering and prediction on the sphere. This has been the subject of a previous report (Colombo, 1979). It must be added that the principles presented here can be generalized to bodies of revolution other than the sphere; to the case where the data are not homogeneous (i.e., a mixture of, say, gravity anomalies, satellite altimetry, etc.); to the case where the coefficients to be estimated are not those of the signal as given, but of some more or less complex linear transformation of this signal, satellite-satellite tracking data being a good example. In fact, the author is at present considering the error analysis of the determination, by least squares collocation, of the potential coefficients from satellite-satellite tracking, following some of the principles of section 2. This will be the subject of another report.

The method developed in section 4 to calculate the covariance between two area means without employing cumbersome numerical integrations is of interest, not only in spherical harmonic analysis, but more generally in filtering, prediction and estimation from mean values on the sphere.

The computer programs described and listed in Appendix B should help the interested reader to implement some of the techniques discussed in this report. The author sincerely hopes that this will be done by workers concerned with improving and further developing such methods.

Above all, the author hopes that he has conveyed to those who had read this far, the idea that the detailed analysis of very large sets of global, regularly sampled data can be done within the computing resources available today to most scientists who work at universities and research institutions everywhere. The processing, be it by numerical quadratures, or by simultaneous adjustment, of "all the data in the world" is not a fanciful thought, but a practical possibility.

References

- Bibby, J., and H. Toutenberg, Prediction and Improved Estimation in Linear Models, John Wiley and Sons, New York, 1977.
- Brigham, E. O., The Fast Fourier Transform, Prentice-Hall, New Jersey, 1974.
- Carslaw, H. S., An Introduction to the Theory of Fourier's Series and Integrals, 3rd. ed., Dover, New York, 1950.
- Colombo, O. L., Optimal Estimation from Data Regularly Sampled on a Sphere with Applications in Geodesy, Department of Geodetic Science Report No. 291, The Ohio State University, Columbus, Ohio, 1979 (a).
- Colombo, O. L., Algorithms for Spherical Harmonic Analysis, paper presented at the Fall meeting of the A.G.U., San Francisco, December 1979 (b).
- Cooley, J.W., and J.W. Tuckey, An Algorithm for Machine Calculation of Complex Fourier Series, Math. Computation, Vol. 19, pp. 297-301, April 1965.
- Gaposchkin, E. M., Averaging on the Surface of a Sphere, Journal of Geophysical Research, Vol. 86, No. B6., June 10, 1980.
- Giacaglia G. E. O. and A. Lundquist, Smithsonian Astrophysical Observatory Special Report No. 344, July 1972.
- Goldstein, J.D., paper presented at the Fall Meeting of the A.G.U., San Francisco, December 1978.
- Hauseholder, A.S., The Theory of Matrices in Numerical Analysis, Dover, New York, 1964.
- Heiskanen, W.A., and H. Moritz, Physical Geodesy, W.H. Freeman, U.S.A., 1967.
- Hobson, E. W., The Theory of Spherical and Ellipsoidal Harmonics, Chelsea, New York, second reprint, 1965.
- Jekell, C., An Investigation of Two Models for the Degree Variances of Global Covariance Functions, Dept. of Geodetic Science Report No. 275, The Ohio State University, Columbus, Ohio, 1978.
- Justice, J.H., A Levinson-Type Algorithm for Two-Dimensional Wiener Filtering Using Bivariate Szegő Polynomials, Proc. IEEE, Vol. 65, No. 6, pp. 882-886, New York, 1977.

Katsambalos, K., The Effect of the Smoothing Operator on Potential Coefficients Determination, Dept. of Geodetic Science Report No. 287, The Ohio State University, Columbus, 1977.

Kaula, W.M., Statistical and Harmonic Analysis of Gravity, J. Geophysical Research, V. 64, pp. 2401-2421, U.S.A., 1959.

Kaula, W.M., Tests and Combinations of Satellite Determinations of the Gravity Field with Gravimetry, Journal of Geophysical Research, vol. 71, 11. 5303-5314, 1966.

Lauritzen, S.L., The Probabilistic Background of Some Statistical Methods in Physical Geodesy, Publication No. 48, Danish Geodetic Institute, Copenhagen.

Lerch, F.J., S.M. Klosko, R.E., Laubscher, C.A. Wagner, Gravity Model Improvement Using Geos-3 (GEM 9 & 10), Journal of Geophysical Research, Vol. 84, No. B8, July 30, 1979.

Lerch, F.J., Gravity Model Development, Earth Survey Applications Division Research Report, 1979, Ed. Lloyd Carpenter, NASA Technical Memo. 80642, Goddard SFC, Greenbelt, Maryland, Jan. 1980.

Lowes, F.J., Comments on 'The Relationship Between Mean Anomaly Block Sizes and Spherical Harmonic Representations' by R.H. Rapp, Dept. of Geophysics and Planetary Physics, University of Newcastle Upon Tyne, U.K., 1978.

Luenberger, D.G., Optimization by Vector Space Methods, Wiley & Son's N.Y., 1969.

Moritz, H., Schwerevorhersage und Ausgleichungsrechnung. Zeitschrift für Vermessungswesen, vol. 90, pp. 181-184, 1965.

Moritz, H., Advanced Least Squares Methods, Dept. of Geodetic Science, Report No. 175, The Ohio State University, Columbus, 1972.

Moritz, H., Covariance Functions in Least Squares Collocation, Dept. of Geodetic Science Report No. 240, The Ohio State University, Columbus, 1976.

Moritz, H., Statistical Foundations of Collocation, Dept. of Geodetic Science Report No. 272, The Ohio State University, Columbus, 1978.

Moritz, H., Advanced Physical Geodesy, Abacus Press, Tunbridge Wells, Kent U.K., and Herbert Wichmann Verlag, Karlsruhe, W. Germany, 1980.

Payne, M.H., Truncation Effects in Geopotential Modelling, Analytical Mechanics Associates, Seabrook, Maryland, 1971.

- Paul, M.K., Recurrence Relationships for the Integrals of the Associate Legendre Functions, Bull. Géodésique, 52, 1978.
- Rapp, R.H., Comparison of Two Methods for the Combination of Satellite and Gravimetric Data, Dept. of Geodetic Science Report 113, The Ohio State University, Columbus, 1968.
- Rapp, R.H., Mean Gravity Anomalies and Sea Surface Heights Derived from Geos-3 Altimeter Data, Dept. of Geodetic Science Report No. 268, The Ohio State University, Columbus, 1977.
- Rapp, R.H., A Global $1^\circ \times 1^\circ$ Anomaly Field Combining Satellite, Geos-3 Altimeter And Terrestrial Anomaly Data, Dept. of Geodetic Science Report No. 278, The Ohio State University, Columbus, 1978.
- Rapp, R.H., Potential Coefficient and Anomaly Degree Variance Modelling Revisited, Depart. of Geodetic Science Report No. 293, The Ohio State University, Columbus, 1979.
- Ricardi, L.J., and J.L. Burrows, A Recurrence Technique for Expanding a Function in Spherical Harmonics, IEEE Transactions on Computers, June 1972.
- Rummel, R., A Model Comparison in Least Squares Collocation, Dept. of Geodetic Science Report No. 238, The Ohio State University, Columbus, Ohio, 1976.
- Rummel, R. and K.P. Schwarz, On the Non-Homogeneity of the Global Covariance Function, Bull. Géodésique, Vol. 51, No. 2, pp. 93-103, Paris, 1977.
- Schwarz, K.P., Application of Collocation: Spherical Harmonics from Satellite Observations, Brosowski and Martensen, vol. 14, pp. 111-132, 1975.
- Sjöberg, L., Potential Coefficient Determinations from 10° Terrestrial Gravity Data by Means of Collocation, Dept. of Geodetic Science Report No. 274, The Ohio State University, Columbus, 1978.
- Tikhonov, A.N. and V.Y. Arsenin, Solutions of Ill-Posed Problems, John Wiley and Sons, N.Y., 1977.
- Trench, W., An Algorithm for the Inversion of Finite Hankel Matrices, S.I.A.M. Journal of Applied Mathematics, Vol. 13, pp. 1102-1107, 1965.
- Uotila, U.A., Useful Matrix Equivalences, Lecture Notes, Dept. of Geodetic Science, The Ohio State University, Columbus, 1978.
- Wagner, C.A. and O.L. Colombo, Gravitational Spectra from Direct Measurements, Journal of Geophysical Research, Vol. 84, No. B9, August 10, 1979.

Appendix A. Term by Term Integration of Formula (4.5), and Rearrangement of Terms in Formula (4.9) to Arrive at Formulas (4.11-a) and (4.14)

The area integrals in expression (4.5) can be split into integrals in θ and λ , and the summation theorem for fully normalized spherical harmonics can be used to replace $P_n(\psi_{pp'})$ with an equivalent expression in the general term of the series:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = & \frac{1}{\Delta_{1j} \Delta_{k1}} \int_{\theta_1}^{\theta_1 + \Delta\theta} \sin \theta d\theta \int_{\theta_k}^{\theta_k + \Delta\theta} \sin \theta' d\theta' \int_{\lambda_1}^{\lambda_1 + \Delta\lambda} d\lambda \int_{\lambda_1}^{\lambda_1 + \Delta\lambda} d\lambda' \sum_{n=0}^{\infty} c_n^{u,v} \cdot \\ & \cdot \sum_{n=0}^n P_{nn}(\cos \theta) P_{nn}(\cos \theta') \cos m(\lambda - \lambda') (2n+1)^{-1} \end{aligned} \quad (\text{A.1})$$

Starting with (A.1), the proof proceeds in four steps, each justifying in turn the taking of one of the four integrals inside the summation symbols. Each time, three theorems are invoked:

The first is the "M-Test" theorem, due to Weierstrass (see, for instance, Carslaw, (1950)):

The series $f(x) = \sum_{n=0}^{\infty} u(x)_n$ will converge uniformly in $a \leq x \leq b$ if there is a convergent series of positive constants $M_0, M_1, \dots, M_n, \dots$ such that, for all x in $a \leq x \leq b$, $|u(x)_n| \leq M_n$ for every positive integer n .

The second theorem is (also according to Carslaw):

If the general term $u(x)_n$ of the series $\sum_{n=0}^{\infty} u(x)_n$ is continuous, and if the series converges uniformly to some function $f(x)$ in the interval $a \leq x \leq b$, then

$$\int_a^b f(x) dx = \int_a^b \sum_{n=0}^{\infty} u(x)_n dx = \sum_{n=0}^{\infty} \int_a^b u(x)_n dx$$

This is a sufficient condition for term by term integration. The third theorem is the mean value theorem for integrals

If $f(x)$ is analytic in $a \leq x \leq b$ then $\int_a^b f(x) dx = (b-a)f(c)$ for some c such that $a \leq c \leq b$.

Proof: The series in the left hand side of (A.1), if all variables but λ' are kept fixed, is uniformly convergent in the interval $\lambda_1 \leq \lambda' \leq \lambda_1 + \Delta\lambda$ because

$$(2n+1)^{-1} c_n^{u,v} \sum_{n=0}^n \bar{P}_{nn}(\cos \theta) \bar{P}_{nn}(\cos \theta') \cos m(\lambda - \lambda') = c_n^{u,v} P_n(\psi_{pp'})$$

and $|c_n^{u,v} P_n(\psi_{pp'})| \leq c_n^{u,v}$ because $\max_{0 \leq \psi \leq \pi} |P_n(\psi)| = 1$ (the argument of P_n is real) for

all n , while $\sum_{n=0}^{\infty} c_n^{u,v} = \text{cov}(u(P), v(P))$ is always finite (equals the value of the "point" covariance when $\psi_{pq} = 0$). The "M-Test" condition is satisfied so the series converges uniformly; therefore, term by term integration with respect to λ' is valid:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) &= \frac{1}{\Delta_{1j}\Delta_{k1}} \int_{\theta_1}^{\theta_1+\Delta\theta} \sin\theta d\theta \int_{\theta_k}^{\theta_k+\Delta\theta} \sin\theta' d\theta' \int_{\lambda_j}^{\lambda_j+\Delta\lambda} d\lambda \cdot \\ &\cdot \sum_{n=0}^{\infty} c_n^{u,v} \sum_{n=0}^n \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \int_{\lambda_1}^{\lambda_1+\Delta\lambda} \cos m(\lambda - \lambda') d\lambda (2n+1)^{-1} \end{aligned}$$

Applying the mean value theorem to the last expression:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) &= \frac{1}{\Delta_{1j}\Delta_{k1}} \int_{\theta_1}^{\theta_1+\Delta\theta} \sin\theta d\theta \int_{\theta_k}^{\theta_k+\Delta\theta} \sin\theta' d\theta' \int_{\lambda_j}^{\lambda_j+\Delta\lambda} d\lambda \cdot \\ &\cdot \sum_{n=0}^{\infty} c_n^{u,v} \sum_{n=0}^n \Delta\lambda \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \cos m(\lambda - \lambda_Q) (2n+1)^{-1} \end{aligned}$$

where $\lambda_1 \leq \lambda_Q \leq \lambda_1 + \Delta\lambda$. Removing the common factor $\Delta\lambda$ from the summation:

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) &= \frac{\Delta\lambda}{\Delta_{1j}\Delta_{k1}} \int_{\theta_1}^{\theta_1+\Delta\theta} \sin\theta d\theta \int_{\theta_k}^{\theta_k+\Delta\theta} \sin\theta' d\theta' \int_{\lambda_j}^{\lambda_j+\Delta\lambda} d\lambda \cdot \\ &\cdot \sum_{n=0}^{\infty} c_n^{u,v} \sum_{n=0}^n \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \cos m(\lambda - \lambda_Q) (2n+1)^{-1} \quad (\text{A.2}) \end{aligned}$$

The general term in the partially integrated series is

$$(2n+1)^{-1} c_n^{u,v} \sum_{n=0}^n \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \cos m(\lambda - \lambda_Q) = c_n^{u,v} P_n(\psi_{pq})$$

where $Q \equiv (\theta', \lambda_Q)$; now $|c_n^{u,v} P_n(\psi_{pq})| \leq c_n^{u,v}$ for all n , and for all λ in the interval $\lambda_j \leq \lambda \leq \lambda_j + \Delta\lambda$, so the "M-Test" is satisfied again and the series is uniformly convergent, and thus integrable, with respect to λ . Therefore

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) &= \frac{\Delta\lambda^2}{\Delta_{1j}\Delta_{k1}} \int_{\theta_1}^{\theta_1+\Delta\theta} \sin\theta d\theta \int_{\theta_k}^{\theta_k+\Delta\theta} \sin\theta' d\theta' \sum_{n=0}^{\infty} c_n^{u,v} \cdot \\ &\cdot \sum_{n=0}^n \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \cos m(\lambda_R - \lambda_Q) (2n+1)^{-1} \quad (\text{A.3}) \end{aligned}$$

where $\lambda_j \leq \lambda_R \leq \lambda_j + \Delta\lambda$. Once more the general term of the twice integrated series satisfies the "M-Test", because

$$|(2n+1)^{-1} c_n^{u,v} \sum_{n=0}^n \bar{P}_{nn}(\cos\theta) \bar{P}_{nn}(\cos\theta') \cos m(\lambda_R - \lambda_Q)| = |c_n^{u,v} P_n(\psi_{pq})| \leq c_n^{u,v}$$

(where $R \equiv (\theta, \lambda_R)$) for all n and, in particular, for θ' in the interval

$\theta_k \leq \theta' \leq \theta_k + \Delta\theta$. In consequence

$$\begin{aligned} \text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) &= \frac{\Delta\lambda^2 \sin \theta_s \Delta\theta}{\Delta_{1j} \Delta_{k1}} \int_{\theta_1}^{\theta_1 + \Delta\theta} \sin \theta \, d\theta \sum_{n=0}^{\infty} c_n^{u,v} \cdot \\ &\cdot \sum_{n=0}^{\infty} \bar{P}_{nn}(\cos \theta) \bar{P}_{nn}(\cos \theta_s) \cos m(\lambda_k - \lambda_0) (2n+1)^{-1} \quad (\text{A.4}) \end{aligned}$$

Finally,

$$|(2n+1)^{-1} c_n^{u,v} \sum_{n=0}^{\infty} \bar{P}_{nn}(\cos \theta) \bar{P}_{nn}(\cos \theta_s) \cos m(\lambda_k - \lambda_0)| \leq |c_n^{u,v} P_n(\psi_{kx})| \leq c_n^{u,v}$$

(where $X \equiv (\theta_s, \lambda_0)$) for all n and, in particular, for θ in the interval $\theta_1 \leq \theta \leq \theta_1 + \Delta\theta$ so the last integral can be put inside of the summations, and the proof of the term by term integration of (4.5) is complete:

$$\text{cov}(\bar{u}_{1j}, \bar{v}_{k1}) = h \sum_{n=0}^{\infty} c_n^{u,v} \sum_{n=0}^{\infty} \bar{P}_{nn}(\cos \theta_r) \bar{P}_{nn}(\cos \theta_s) \cos m(\lambda_k - \lambda_0) \quad (\text{A.5})$$

where $\theta_1 \leq \theta_r \leq \theta_1 + \Delta\theta$ and $h = \Delta\lambda^2 \sin \theta_s \Delta\theta \sin \theta_r \Delta\theta / \Delta_{1j} \Delta_{k1}$

The general term in (A.5) satisfies the "M-Test":

$$|z_n| = |(2n+1)^{-1} c_n^{u,v} \sum_{n=0}^{\infty} \bar{P}_{nn}(\cos \theta_r) \bar{P}_{nn}(\cos \theta_s) \cos m(\lambda_k - \lambda_0)| \leq c_n^{u,v}$$

Since the series $\sum_{n=0}^{\infty} c_n^{u,v}$ converges, any series of positive terms $|z_n|$ satisfying $|z_n| \leq c_n$ must converge also: the "M-Test" condition implies the absolute convergence of $\sum_{n=0}^{\infty} z_n$. Absolutely convergent series can have the order of their terms changed arbitrarily, without changing the value of their limit sums. This justifies the reordering of (4.9) that leads to expression (4.11-a).

Appendix B: Computer Programs (Descriptions and Listings)

This Appendix contains the description and listing of each of the major subroutines developed in the course of this research, together with their own auxiliary routines. The listings of the main subroutines contain explanatory comments that the author hopes will be of help to those who may use them. The description accompanying each main program defines arguments, gives the dimensions of the memory arrays required, mentions the relevant formulas from preceding sections, and gives a brief explanation of the various segments of the program.

B.1 General Programming Considerations

Only one-dimensional arrays are used in the software described here. The language used for all of them is FORTRAN IV ; some subroutines from the International Mathematical and Statistical Libraries Inc. (IMSL) are called by the main subroutines. Implicit DO loops in READ or WRITE statements have been avoided as much as possible, because their execution may be rather slow, depending on the compiler; instead, subroutines FREAD, FWRITE and REWIND are used for all input/output operations involving large files on tape or disk, in some of the subroutines. All operations involving real arithmetic have been coded in double precision (8 bytes, or 32 bits), which is equivalent to retaining the first 7 significant figures in all arithmetic operations.

The arrays containing the associated Legendre functions or their integrals, as the case may be, are arranged first by degree, and then by order : 00, 10, 11, 20, 21, 22, . . . (N_{max} , N_{max}). The \bar{C}_{nn}^m are arranged accordingly, always in two separated arrays: one for the \bar{C}_{nn}^0 and another for the \bar{C}_{nn}^1 . In order to get the value of the element 'nm' from one of this arrays, the following formula is used:

$$k = \frac{1}{2} n (n + 1) + m + 1$$

where k is the position of this element in the one-dimensional array. When the elements are recovered sequentially from the beginning (00), the following type of DO loop is used:

```
KOUNT = 1
DO XX      N1 = 1, Nmax
DO XX      M1 = N1, Nmax
```

```
LEGEND (KOUNT) = ARRAY (KOUNT) **2
```

```
XX KOUNT = KOUNT + 1
```

where, in this particular example, the nm (n=N1-1, m=M1-1) element in array L LEGEND is equated to the square of the nm element in ARRAY. Avoidance of two-dimensional memory arrays results in considerable improvements in efficiency.

B.2 Subroutine SSYNTH

This subroutine computes the sum of a spherical harmonic series complete to degree and order NMAX at each one of the $2N^2$ points or blocks in an equal angular grid. The subroutine can calculate point values (IFLAG = 0) or area means (IFLAG = 1). The number of rows or parallels N (Nyquist frequency) must be even. Subroutines FFTP from the IMSL Double Precision Library is used to calculate the sum of the series along rows by means of the Mix Radix Fast Fourier Transform algorithm.

The procedure used is that described in paragraphs (1.6) and (1.7). The symmetry of the grid with respect to the equator, and the corresponding even-odd symmetry of the values of the Legendre functions or their integrals, (i.e., the χ_i^{aa} of section 1) are exploited. The values of those functions, or of their integrals, are read from mass storage (disk or tape) into array ROW, in the order described in the previous paragraph. All the values for one latitude, or "row," are read at once, so the dimension of ROW is $\frac{1}{2}(N_{max} + 1)(N_{max} + 2)$. All the coefficients C_{nm}^α are also stored in core, the corresponding RCNM and RSNM arrays (for \bar{C}_{nm} and \bar{S}_{nm} respectively) have the same dimension (the \bar{S}_{nm} are included, though they are all zero). The output consists of $2N^2$ values in array DATA. This array is organized in rows, from North Pole to South Pole. The rows, of $2N$ points or blocks each, have their values written consecutively. The following is the list of arrays, and their dimensions:

NAME	DIMENSION	TYPE
ROW	$RD = \frac{1}{2}(N_{max} + 1)(N_{max} + 2)$	REAL * 8
RCNM	RD	"
RSNM	RD	"
X	RD	INTEGER * 2
DATA	$2N^2$ (N = 180/BLOCK)	REAL * 8
CR1	N + 1	"
CR2	N + 1	"
SR1	N + 1	"
SR2	N + 1	"
AM	N + 1	"
BM	N + 1	"
F AUX1	4N	"
F AUX2	4N	"
F IWK	see IMSL Handbook	INTEGER * 4
F LL	" " "	LOGICAL * 4
F A	" " "	REAL * 8
IV	N + 1	"

("F" designates those arrays required by the IMSL subroutine FFTP). In addition, the size (in degrees) of the blocks is defined by BLOCK; $NPP = \frac{1}{2}(N_{max} + 1)(N_{max} + 2)$; IU is the number of the unit (disk, tape) from which the Legendre functions or their integrals are to be read.

Array X contains information on whether a given χ_i^{aa} is even or odd;


```

0029 ENH 1) = 0.00
0030 GO TO 8
0031 DO 6 NI = 1, NP1
0032 R = NI - 1
0033 BL2H = BL2H + NI
0034 AMH1) = DCOS(BL2H)
0035 6 ENH1) = -DSIN(BL2H)
0036 8 RHT = RC + NC2
0037 DO 9 I = 1, NP1
0038 9 IV(I) = IS(I - 1)/2
0039 ISM = -RC
0040 ISM = RHT
0041 PH11 = 90.00 + 2*PI*RC
0042 PH12 = PH11 - BLK
C C C C
0043 DO 40 NR = 1, NR02
0044 FROMIR = 0.300
0045 IF (IFLAG.NE.1) GO TO 12
C C C C
0046 COMPUTING THE INVERSE OF THE AREA OF A BLOCK IN
0047 ROW NR
0048 FROMIR = 0.300/((DSIN(PH11) - DSIN(PH12))*BLK)
0049 PH11 = PH11 - BLK
0050 PH12 = PH12 - BLK
12 ISM = ISM + RC
ISM = ISM - RC
C C C C C C C C
0051 FORM THE COS AND SIN FOURIER COEFFICIENTS OF EACH "ROW"
0052 UP TO FROMIR * MAX. IN ARRAYS CRI, SRI AND
0053 CRI, SRI, RESPECTIVELY.
0054 (INNER LOOP)
0055 READ THE (INTEGRALS OF) NORMALIZED LEGENDRE FUNCTIONS
0056 CORRESPONDING TO THE NORTHERN "ROW", FROM UNIT "10"
0057 (UNFORMATTED RECORDS)
0058 READ(10) ROW
0059 DO 20 NI = 1, NP1
0060 C1 = 0.00
0061 C2 = 0.00
0062 S1 = 0.00
0063 S2 = 0.00
C C C C
0064 THIS IS THE BEGINNING OF THE INNERMOST DO LOOP
0065 DO 15 NI = NI, NP1
0066 NUC = IV(NI) * NI
0067 R = ROW(NUC)
0068 S = R * SIN(NUC)
0069 C = R * COS(NUC)
0070 IF (X(RUC) .EQ. -1) GO TO 14
0071 C1 = C1 + C
0072

```

```

0044 C2 = C2+C
0045 S1 = S1+S
0046 S2 = S2+S
0047 GO TO 16
0048 14 C1 = C1+C
0049 C2 = C2+C
0050 S1 = S1+S
0051 S2 = S2+S
0052 16 CONTINUE
0053
0054
0055
0056
0057
0058
0059
0060
0061
0062
0063
0064
0065
0066
0067
0068
0069
0070
0071
0072
0073
0074
0075
0076
0077
0078
0079
0080
0081
0082
0083
0084
0085
0086
0087
0088
0089
0090
0091
0092
0093
0094
0095
0096
0097
0098
0099
0100
0101
0102
0103
0104
0105
0106
0107
0108
0109
0110
0111

```

END OF INNERMOST LOOP

CR1(N) = C1*AM(N)-S1*BM(N)
CR2(N) = C2*AM(N)-S2*BM(N)
SR1(N) = S1*AM(N)+C1*BM(N)
SR2(N) = S2*AM(N)+C2*BM(N)

20 CONTINUE
CR1(1) = 2.00*CR1(1)
CR2(1) = CR2(1)*2.00

"ALIAS" THE COEFFICIENTS

DO 25 N1 = 1, NC2P
NA = N1
NB = -N1+2
DO 25 K = 1, NP1
NA = NA+NC
NB = NB+NC
IF(NB.GT.NP1) GO TO 25
IF(NA.GT.NP1) GO TO 21
CR1(N) = CR1(N)+CR1(NA)+CR1(NB)
CR2(N) = CR2(N)+CR2(NA)+CR2(NB)
SR1(N) = SR1(N)+SR1(NA)-SR1(NB)
SR2(N) = SR2(N)+SR2(NA)-SR2(NB)
GO TO 25
21 CR1(N) = CR1(N)+CR1(NB)
CR2(N) = CR2(N)+CR2(NB)
SR1(N) = SR1(N)-SR1(NB)
SR2(N) = SR2(N)-SR2(NB)
25 CONTINUE

FORM THE COMPLEX FOURIER COEFFICIENTS

K1 = -1
K2 = 0
DO 30 N1 = 1, NC
K1 = K1+2
K2 = K2+2
IF(N1.GT.NC2P) GO TO 26
AUX1(K1) = CR1(N1)
AUX2(K1) = CR2(N1)
AUX1(K2) = -SR1(N1)
AUX2(K2) = -SR2(N1)
GO TO 30
26 NV = NP-N1+1
AUX1(K1) = CR1(NV)
AUX2(K1) = CR2(NV)

PAGE 0004

12/29/84

DATE - 00242

001177

FORTRAN IV C1 MESSAGE 2.0

0112 AUX1(K1) = PR1(MO)
0113 AUX2(K1) = SR2(MO)
0114 30 CONTINUE

CCCCC

OBTAIN THE CONTENTS OF EACH ROW IN 'DATA' AS FOURIER
TRANSFORMS OF 'AUX1' AND 'AUX2'.

0115 CALL FFTP(AUX1, MC, IWE, A, LL)
0116 CALL FFTP(AUX2, MC, IWE, A, LL)
0117 K1 = -1

CCCCC

END OF OUTER AND INNER LOOPS.

0118 DO 40 I = 1, MC
0119 K1 = K1+2
0120 DATA(150+1) = AUX1(K1)*FRONTIN
0121 DATA(150+1) = AUX2(K1)*FRONTIN
0122 40 CONTINUE
0123 RETURN
0124
0125 END

40

arrays CR1, SR1, and CR2, SR2, respectively, contain the Fourier coefficients $\alpha_n^i(a_n^i)$ and $\beta_n^i(b_n^i)$ of rows 1 and N-1-1; arrays AM and BM contain the values of A(m) and B(m), as defined in expression (1.7); the auxiliary array IV contains the numbers $2n(n+1)$ needed to locate individual elements within arrays ROW, RCNM, and RSNM, when they are not addressed sequentially.

The comments in the listing are probably enough to understand most of it on close inspection; one point however may be worth explaining further: the "aliasing" of the Fourier coefficients has been incorporated to take care of the case when $N_{max} > N-1$. In such situation the $\alpha_n^i(a_n^i)$ and $\beta_n^i(b_n^i)$ become aliased, as Fourier coefficients must, and it is their aliased values that the FFT subroutine requires to compute the values of the spherical harmonic expansion along parallels. The formulas for the aliased coefficients are

$$\hat{\alpha}_n^i = \alpha_n^i + \sum_{k=1}^{KM} (\alpha_{n+KM}^i + \alpha_{n-KM}^i) \quad (B1-a)$$

$$\hat{\beta}_n^i = \beta_n^i + \sum_{k=1}^{KM} (\beta_{n+KM}^i - \beta_{n-KM}^i) \quad (B2-b)$$

where KM is a large enough integer. A similar expression applies to a_n and to b_n .

The arrangement of the output in latitude corresponds, in the case of area means, to the intervals on which the \bar{P}_{nn} are integrated; for point values, it is defined by the latitudes θ_i at which the $\bar{P}_{nn}(\cos \theta_i)$ have been precomputed. As regards longitude, the grid starts from the zero meridian used for defining the coefficients. In the case of point values it is usual to compute all values at the center of each block. To do this, the \bar{P}_{nn} must be precomputed at the latitudes of the center points, while the longitudes are taken care by modifying the coefficients as follows

$$\bar{C}'_{nn} = \bar{C}_{nn} \cos m \frac{\Delta\lambda}{2} + \bar{S}_{nn} \sin m \frac{\Delta\lambda}{2} \quad (B2-a)$$

$$\bar{S}'_{nn} = \bar{S}_{nn} \cos m \frac{\Delta\lambda}{2} - \bar{C}_{nn} \sin m \frac{\Delta\lambda}{2} \quad (B2-b)$$

This is equivalent to rotating the grid eastwards from the zero meridian by $\frac{\Delta\lambda}{2}$.

B.3 Subroutine HARMIN

This subroutine implements either the algorithm of paragraph (1.5) for the harmonic analysis of area means, or that of paragraph (1.7) for the analysis of point values.

The subroutine calls IMSL's FFCSIN to calculate the a_n^i , b_n^i or the α_n^i , β_n^i by means of the Fast Fourier Transform (Mix Radix) algorithm. It also calls subroutine QUADFS, that returns in array A the de-smoothing

00010700
00010800
00010900

```

6 BK(N1) = -DSIN(RL2ND)
7 CALL QDADFS(A,N1,BLOCK,NC)
DO 8 I = 1,RDB
  BCHK(I) = 0.00
8 BCHK(I) = 0.00
DO 9 I = 1,N1
  IF(IPLAG.EQ.1) A(I) = A(I)*PT
9 IV(I) = 1*(I-1)/2

```

BEGINNING OF MAIN (OUTER LOOP, WHERE DATA "NOMS" (I.E.,
CONSTANT LONGITUDE BLOCKS) ARE ANALYZED TWO AT THE TIME
TO EXPLOIT EQUATORIAL SYMMETRY.

```

10 AUX1(I) = DATA(ISH+1)
CALL FPCSIN(AUX1,NC,CRI,SRI,IWO)
CALL FPCSIN(AUX2,NC,CR2,SR2,IWO)
UNIT = 10* (INTEGRALS OF) LEGENDRE FUNCTIONS FROM
      UNIFORM RECORDS)
READ(IU) NOV

```

UNIT = 10* (INTEGRALS OF) LEGENDRE FUNCTIONS FROM
UNIFORM RECORDS)

BEGINNING OF INNER LOOP

```

DO 20 N1 = 1,N1
  N1 = N1-1
  IF(N1.LE.NC2) GO TO 12
  N1 = NC-N1
  SRI(N1) = -SRI(N1)
  SR2(N1) = -SR2(N1)
12 CONTINUE
  C1 = CRI(N1)*AK(N1)+SRI(N1)*BK(N1)
  C2 = CR2(N1)*AK(N1)+SR2(N1)*BK(N1)
  S1 = SRI(N1)*AK(N1)-CRI(N1)*BK(N1)
  S2 = SR2(N1)*AK(N1)-CR2(N1)*BK(N1)
  CP = C1+C2
  CM = C1-C2
  SP = S1+S2
  SM = S1-S2
  IF(IPLAG.EQ.1) GO TO 17
  CP = CP*FINN
  CM = CM*FINN
  SP = SP*FINN
  SM = SM*FINN
  INNERMOST DO LOOP

```

```

17 DO 20 N1 = 1,N1
  NUC = IV(N1)+N1
  R = ROM(NUC)
  IF(X(NUC).EQ.-1) GO TO 18

```

-109-

PORTMAN IV C1 RELEASE 2.0 QUADFS DATE = 00242 12/29/34 PAGE 0001
 SUBROUTINE QUADFS(A, NN1, BLOCK, NC)
 C
 C THIS SUBROUTINE COMPUTES THE VECTOR OF INTEGRATION WEIGHTS.
 0001
 0002 IMPLICIT REAL*8(A-H,O-Z)
 0003 DIMENSION A(1), POLS(30)
 0004 COMMON/AN/ ANH(30), A2NN1(30)
 0005 NN = NN1
 0006 P = 3.141592653589793D0
 0007 BROOV = P/100.D0
 0008 NC2 = NC/2
 0009 NC3 = NC/3
 0010 N1 = BLOCK*BROOV
 0011 BLICK = BROVT/BS(1, BLK)*BLK/P1/BROOV
 0012 F = 1.D0/(1.D0 - DCS(BLICK*BROOV))
 0013 DO 5 N = 2, NN1
 0014 ANH(N) = -(N-1)*1.D0/N
 0015 A2NN1(N) = (2.D0*N-1.D0)/N
 0016 CALL LEQPOL(POLS, BLICK, NN1)
 0017 A(1) = 0.D0
 0018 A(2) = 0.D0
 0019 DO 10 N = 2, NN
 0020 A(N+1) = 1.D0/(F*(2*N+1)*POLS(N-1)-POLS(N+1))
 0021 IF(N.LE.NC3) A(N+1) = A(N+1)*2
 0022 IF(N.GT.NC3) A(N+1) = 1.D0
 0023 10 CONTINUE
 0024 WRITE(6,170) (1,A(1),1=1,NN1)
 0025 170 FORMAT(1X,4(2X,13.2X,C20.10))
 0026 RETURN
 0027 END

```

0001
0002
0003
0004
0005
0006
0007
0008
0009
0010
0011
0012
0013
0014
0015
0016
0017

FORTRAN IV C1 RELEASE 2.0
SUBROUTINE LECFOL(POLS,PSI,NMAX)
IMPLICIT REAL*8(A-H,O-Z)
COMMON /AN/ ANH(301),A2NH(301)
DIMENSION POLS(1)
PI = 3.141592653589793D0
DRCOV = PI/180.D0
10 T = DCON(PHI*DRCONV)
PHI = 1.D0
PR = T
POLS(1) = T
DO 20 N = 2,NMAX
POLS(N) = ANH(N)*PHI+A2NH(N)*PR*T
PHI = PH
PR = POLS(N)
20 CONTINUE
RETURN
END

PAGE 0001
00026100
00026200
00026300
00026400
00026500
00026600
00026700
00026800
00026900
00027000
00027100
00027200
00027300
00027400
00027500
00027600

```

factors μ_n . QUADFS calls LEGPOL, a subroutine that computes the Legendre polynomials up to degree $NN + 1$ needed for the β_n^{-1} in μ_n .

The data is arranged as in SSYNTH, in array DATA, before the subroutine is called. Afterwards, the contents of DATA are destroyed, as the a_n^1 , b_n^1 or the α_n^1 , β_n^1 are formed in place of them, row by row, by FFCSIN. The resulting coefficients' estimates are put into arrays RCNM and RSNM, in the same order as for HARMIN. The other arrays, with the exception of A, are as in SSYNTH. The same is true of the scalars, with the exception of NN. NN is the highest degree and order to be estimated. NDD is the total number of Legendre functions, or their integrals, to be read from unit IU, per G_1 . This number is $(NN+1) + (NN+2)/2$. A is a REAL*8 array of dimension $NN + 1$. The dimension in QUADFS allows for a maximum $NN = 300$; for larger solutions, the dimensions there and in LEGPOL must be increased accordingly.

In the case of point data, the estimated coefficients are computed using a center point formula that assumes that the data are situated at the centers of the blocks; the resulting coefficients are referred, nonetheless, to a grid starting at the zero meridian (the "rotation" of the coefficients takes place between statements 0071 and 0073, when $IFLAG = 0$). When $IFLAG = 1$, the area means formula (1.30) is computed; the $\chi_i^n = \mu_n \int_0^\pi \sin \theta d\theta$, and the μ_n are those produced by QUADFS, as already mentioned. The integrals of the Legendre functions are read from unit "IU", as in SSYNTH (same format), and the size of the blocks is specified by BLOCK (in degrees). The version of QUADFS listed here implements the "composite" estimator of paragraph (3.3). If another is desired, this can be achieved simply by replacing lines 0021 through 0024 in QUADFS.

B.4 Subroutine NORMAL

This subroutine creates the optimal estimators for the \bar{C}_{nn}^α based on the formation and inversion of the $R(m)$ matrices described in paragraph (2.10). The algorithm exploits the fact that $(C_{zz} + D)$ is a block matrix of Toeplitz circulant sub-matrices. This subroutine is meant only for mean values.

The grid is as in SSYNTH and HARMIN. The symmetry with respect to the equator is only partly exploited; matrix D may not be persymmetric, so the total matrix $(C_{zz} + D)$ may not be so either. C_{zz} however, is always persymmetric, and this is taken into account to save computing and storage. A general diagonal matrix D corresponds to a rather broad class of actual problems, such as the analysis of the $5^\circ \times 5^\circ$ real gravity anomalies described in paragraph (3.4).

This subroutine requires four input/output units: 8 (read only) contains the values of the integrals of the Legendre functions, row by row, arranged as in SSYNTH or HARMIN; 10 contains the right hand sides of the "reduced normals" $\underline{k}_n = R(m) \underline{\chi}^n$ (expression (2.58)); 15 contains the $R(m)$ matrices, ordered by

READ THE (INTEGRALS OF) NORMALIZED LEGENDRE FUNCTIONS
FOR ROTATED NOV

SUBROUTINE NORMAL(RMAX, NR, DCRID, NEZUL, ICEO, NRUN, ROMP, RNOQ, RES, S, A, 00000100
2 UL, NCS, V, DVAR, FC, PH, RS)

THIS SUBROUTINE IMPLEMENTS THE FAST COLLOCATION
ALGORITHM. IT CREATES THE OPTIMAL FILTER FOR ESTIMATING
SPHERICAL HARMONIC COEFFICIENTS OF AREA MEANS (EQUAL
ANGULAR) ON A SPHERE. A CHOICE CAN BE MADE OF EITHER
GEOCENTRIC OR GEODETIC COORDINATES.

THE GRID MUST BE BOTH COMPLETE AND SYMMETRICAL RESPECT
TO THE EQUATOR
BECAUSE OF THE GRID'S SYMMETRY, THE "SIGNAL" PART OF EACH
R(N) MATRIX IS PERSYMMETRIC, AND FULL ADVANTAGE IS TAKEN
OF THIS FACT WHEN SETTING UP THE R(M)'S. THE "NOISE" PART,
ON THE OTHER HAND, CAN BE ANY DIAGONAL MATRIX (ASYMMETRIC).

INTEGRALS INMP ARE READ IN FROM UNIT 8, STORED, AFTER
MULTIPLYING THEM BY CERTAIN FACTORS, IN UNIT 10.
THE R(M)'S MATRICES ARE STORED IN UNIT 15.
OPTIMAL "QUADRATURES" WEIGHTS" IN UNIT 30.

THIS PROGRAM USES SUBROUTINES "LUDECT" AND "LUZLMP" FROM
THE IMSL DOUBLE LIBRARY. IN TURN, THESE USE OTHER
SUBROUTINES FROM THE IMSL SINGLE LIBRARY. THIS PROGRAM
ALSO CALLS SUBROUTINE "YUR" TO COMPUTE THE FOURIER
COEFFICIENTS OF THE BLOCK COVARIANCES, AND SUBS "YREAD",
"YWRITE", AND "REWIN" FOR FAST INPUT-OUTPUT. THE USER
CAN REPLACE THESE WITH HIS OWN OR WITH THE STANDARD "READ",
"WRITE", AND "REWIND" COMMANDS. RPPD, RBL, ETC. ARE THE
LENGTHS (I.E., IN BYTES) OF THE ARRAYS TO BE READ
OR WRITTEN.
IF THE R(M) MATRIX OF A PARTICULAR ORDER CANNOT BE
INVERTED, NULL QUADRATURES "WEIGHTS" FOR THAT ORDER ARE
STORED IN UNIT 30.

PROGRAMMED BY OSCAR L. CULORNO, GEOD. SC., OHIO STATE U., SEPT. 1979.

IMPLICIT REAL*8(A-H,O-Z)
DIMENSION INMP(1), RNOQ(1), RES(1), S(1), A(1), UL(1)
DIMENSION ANPQ(400), ES(1), R(200), X(200), YF(400),
2 I(400), XQ(200), BT(200), FC(1), PH(1), PH(200), DVAR(1), FINP

3 (400)
COMMON/NO/ RT(310)

BASIC CONSTANTS

PI = 3.14159265358979300
DLCOV = PI/110.00
F = 1.00/298.25700

PROBLEM DEFINITION

```

0009      BLOCK SIZE IN DEGREES, MINIMUM. 1 DEGREE
0010      TO GEOMETRIC LATITUDES ARE TO BE CONVERTED
0011      MAX. DEG. AND ORDER TO BE ANALYZED (LESS THAN
0012      NO. OF COLUMNS, ALWAYS EVEN)
0013      NO. OF ROWS. THERE IS NO EXHAUSTIVE ROW.
0014      MAX. DEGREE IN TRUNCATED EXPANSION OF BLOCK CO-
0015      VARIANCE FUNCTIONS, NOT GREATER THAN NC
0016      UNIT 16.
0017      IF N(N)'S ARE READ FROM UNIT 18
0018      OF BEING CREATED.
0019      REGULARIZATION PARAMETER, ADDED TO ALL DIAGONAL
0020      TERMS OF THE N(N) MATRICES.
0021      FOR = DCRID=DCRCONV
0022      NC = 240.00/DCRCONV+0.00000100
0023      NC2 = NC+0.25+0.00000100
0024      NR = NC2
0025      NR1 = NR+1
0026      NP = 1
0027      NP1 = NP+1
0028      NP2 = (NP1*(NP1+1))/2
0029      NLL = (NR1*(NR1+1))/2
0030      NLLB = 8*NPP
0031      NLLB = 8*NLL
0032      NC21 = NC2+1
0033      NC2B = 8*NC2
0034      DVARS = 0.10
0035      DO 3 I = 1, NC
0036      FC(I) = 0.10
0037      PR(I) = 0.10
0038      X0(I) = 0.10
0039
0040      FORM ARRAY OF GEOMETRIC OR GEOMETRIC GRID'S GEOMETRIC LATITUDES
0041      FM = F/(2.00-F)
0042      KM = 2.00+FM/(1.00+FM**2)
0043      KC2P = KC2+1
0044      ARG = 90.00
0045      DO 6 I = 1, NC2P
0046      PH(I) = ARG-DCRCONV
0047      IF(GEO.EQ.2) PH(I) = PH(I)-EXPDSIN(2.00*PH(I))
0048      ARG = ARG-INCRIID
0049      DO 15 I1 = 1, NP1
0050      I = I1-1
0051      FINTP(I1) = DSQRT(DVAR(I)/(2.00+I+1))
0052      FINTP(I) = DSQRT(DVAR0)
0053
0054      READ TAPE OF INTEGRALS OF LEGENDRE FUNCTIONS FROM UNIT 8
0055      MULTIPLY THEN BY THE "ROW FACTORS" (INVERSE OF BLOCK AREA,
0056      AND BY SQUARE ROOT OF DEGREE VARIANCE (PER COEFF.), STUNTING
0057      THESE IN UNIT 16
0058
0059      INB = -NR1

```

-115-

PAGE 0004

12/29/34

DATE = 00242

NORMAL

RELEASE 2.0

PORTMAN IV G1

00016400
00016500
00017700
00017900
00018000

00018100
00018200
00018300
00018400

00018500
00018700
00018800

00019200

00019600
00019700
00019800
00019900
00020000
00020100
00020200
00020300
00020400
00020500
00020600
00020700
00020800
00020900
00021000
00021100
00021200
00021300
00021400
00021500
00021600
00021700
00021800
00021900
00022000
00022100
00022200
00022300
00022400
00022500

DO 103
103 ROW(1) = ROW(1)
G
G
G
G
OBTAIN THE FOURIER COEFFICIENTS OF THE BLOCK COVARIANCES
CORRESPONDING TO ROWS P AND Q IN CDD.
AND ALIAS THEM TO FORM THE ELEMENTS OF THE R(M) MATRICES

105 CONTINUE
DO 111
111 CONTINUE
N = N-1
N = N-1
IF(N.EQ.0 OR N.EQ.NC2) N = NC
CALL FUNK(AN, NI, ROWP, ROWQ, IDO, FF, NP1)
ANPQ(NI) = AN
N1 = NC-N1
N2 = NC-N1
112 CONTINUE
IF(N1.GT.NP1) GO TO 111
CALL FUNK(AN1, NI, ROWP, ROWQ, IDO, FF, NP1)
ANPQ(N1) = ANPQ(N1)+AN1
IF(N2.GT.NP1) GO TO 105
CALL FUNK(AN2, N2, ROWP, ROWQ, IDO, FF, NP1)
ANPQ(N2) = ANPQ(N1)+AN2
106 N1 = N1+NC
N2 = N2+NC
GO TO 112
111 ANPQ(N1) = ANPQ(N1)+N
IS = IS+NN1
DO 114
114 S(IS+NL) = ANPQ(NL)
NL = 1, NN1
115 CONTINUE
120 CONTINUE
STORE THE LOWER DIAGONAL PART OF R(M) COLUMN BY COLUMN
ON DISK FILE ('A1T 15').

DO 116
116 IS = -NN1
DO 115
115 IS = IS+NN1
115 S(1) = S(18+NC2)
CALL FWRITE(S(1), 18, NC28, 8999, 8999)
116 CONTINUE
CALL REVINDM(10)
NP = NP+1
IF(NP.LE.NC4) GO TO 1000
1000 CONTINUE
ADD "REGUL" (THE REGULARIZATION CONSTANT) TO DATA "ROW
VARIANCES"
DO 217
217 W(1) = W(1)+REGUL
CONTINUE

0009
0000

0001
0002
0003
0004
0005
0006
0007
0008
0009
0010
0011
0012
0013
0014
0015
0016
0017
0018

0119
0120
0121
0122
0123
0124
0125
0126
0127
0128
0129

0130
0131
0132

PAGE 0003

12/29/34

DATE - 80242

TV-14

FAST 2.0

00022400
00022100
00022600
00022700
00023400
00023400
00023400
00021100
00021200
00023100
00023400
00023100
00023100
00023700
00023900
00024000
00024100
00024200
00024300
00024400
00024500
00024600
00024700
00024900
00023100
00023200

BEGIN MAIN LOOP WHERE ERRORS ARE ESTIMATED.

66

```
DO 300 MI = 1,NR1
CALL REVIND(15)
FFF = FFF(MI)*MC
M = MI-1
IF(N.EQ.0) GO TO 214
DO 213 I = 1,M
213 CALL FREAD(SB(1),15,MC288
```

2

● ● ● ● ●

READ IN MATRIX R(ND FROM UNIT 15. COLUMN BY COLUMN,
AND PUT IT IN SYNETRAC STORAGE FORM IN VECTOR A.

20

FORM THE CHOLESKY FACTOR UL.

www

```
CALL LUDCT(A,UL,M2,D1,E2,IERD)
IF(IER.EQ.129) WRITE(6,221) N
221 FORMAT(/ / * * * MATRIX R(M)+V
      2SINGULAR * * */)
IF(IER.EQ.129) GO TO 298
```

2

010 010 010 010

SOLVE THE REDUCED EQUATIONS $R(M, X(M, N)) = B(M, N)$ FOR ALL
 $M = 1, \dots, M$ AND ALL $N = 1, \dots, N$, AND SAVE THE OPTIMAL "QUADRATURE WEIGHTS".
 OR SOLUTIONS, IN UNIT 30.

33

FIRST FORM BK N. 4)

```

DO 240      NI = NI, NI
      SNH = 1. DO
      IF((NI-NI)/2)*2.WE.(NI-NI) SNH = -1. DO
      ISH = -NLL
      KT = IDB(NI)+NI
      DO 230      I = 1, NC4
      ISH = ISH+NLL
      NI = NBT(ISH-KT)*FIMF(NI)
      NC2 = NC2-1, NC2-1
      NC2(1) = B(1)
      IF(SNH.EQ.-1) B(NC2(1)) = -B(1)
230

```

2

B9106
29106
59106
49106
39106
19106

12/29/34

DATE = 00242

NORMAL

RELEASE 2.0

SAVE RESULT IN UNIT 30 .

CALL LUELF(UL,B,MC2,X)

COMPUTE THE A POSTERIORI STANDARD ERRORS OF THE COEFFICIENTS
PER DEGREE.

PROISE = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

PCSD = 0.50

CHECK FOR AGREEMENT BETWEEN ORIGINAL AND RECONSTRUCTED R.E.S.'S
BY TESTING THE (R,N) CASES.

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

90 = 0.50

FINISH COMPUTING THE A POSTERIORI COEFFICIENTS'
ERRORS PER DEGREE.

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

90 310 I = 2.00

SUBROUTINE ANALYS(X,BLOCK,DATA,RE,NR,NR,NR,AM,BM,NRONS,
2 NOW,CR,SR,IRK,CAA,CBB,SAA,SBH)

THIS SUBROUTINE USES THE OUTPUT OF SUBR. "NORMAL".
READ FROM UNIT 10, TO COMPUTE THE SPHERICAL HARMONIC
COEFFICIENTS OF THE DATA STORED IN ARRAY "DATA" AS
(BLOCK) DECREASES X (BLOCK) DECREASES AREA REARS. USING
AN ALGORITHM FOR FAST COLLOCATION. THIS SUBR. CALLS SUBR.
"FFCSIN" FROM THE "INSLIM" LIBRARY.
THE ORIGINAL CONTENTS OF ARRAY "DATA" ARE DESTROYED.
THE GRID IS SUPPOSED TO BE EQUAL ANGULAR.
THE NUMBER OF BLOCKS PER NOW (NC) MUST BE EVEN.

PROGRAMMED BY OSCAR L. COLOMBO, DEPT. OF GEODET. SC., OHIO S.U.

IMPLICIT REAL*8(A-M,O-Z)
DIMENSION NRNM(1),NRNM(1),DATA(1),AM(1),BM(1),X(1),FR(1),CR(1)
2, SR(1),IRK(1),CAA(1),CBB(1),SAA(1),SBH(1)

DIMENSION NOW(1)
PI = 3.141592653589793D0
BROCONV = PI/180.D0
BLK = BLOCK*DMCONV

NR1 = NR+1
NR0 = NR1*(NR1+1)/2
NC2 = NC/2
NCP = NC/2

NC2P = NC2+1
NR0 = NR0S+8
DO 5, M1 = 2, NR1

NR = NR1-1
AM(N1) = DSIN(M*BLK)/M
BM(N1) = (1.D0-DCOS(M*BLK))/M

AM(1) = BLK
BM(1) = 0.D0
DO 6, M1 = 1, NR0

NRNM(1) = 0.D0
NRNM(1) = 0.D0
NRNM(1) = 0.D0

ISH = -NC
DO 15, NR = 1, NRONS
ISH = ISH+NC
NR1P = ISH+NC

DO 10, I = 1, NC
NR1P = ISH+NC
CALL FFCIN(NRNM,NC,CR,SR,IRK)

NR1P = ISH+NC
DO 15, NR = 1, NRONS
NR1P = ISH+NC

DO 15, NR = 1, NRONS
NR1P = ISH+NC
NR1P = ISH+NC

DO 15, NR = 1, NRONS
NR1P = ISH+NC
NR1P = ISH+NC

DO 15, NR = 1, NRONS
NR1P = ISH+NC
NR1P = ISH+NC

DO 15, NR = 1, NRONS
NR1P = ISH+NC
NR1P = ISH+NC

CCCCCCCCCCCCCCCC

0002
0003
0004
0005
0006
0007
0008
0009
0010
0011
0012
0013
0014
0015
0016
0017
0018
0019
0020
0021
0022
0023
0024
0025
0026
0027
0028
0029
0030
0031
0032
0033
0034
0035
0036
0037
0038
0039
0040
0041
0042
0043

00056100
00056300
00056400
00056500
00056600
00056700
00056800
00056900
00057000
00057100
00057200
00057300
00057400
00057500
00057600
00057700
00057800
00057900
00058000
00058100
00058200
00058300
00058400
00058500
00058600
00058700
00058800
00058900
00059000
00059100
00059200
00059300
00059400
00059500
00059600
00059700
00059800
00059900
00060000
00060100
00060200
00060300
00060400
00060500
00060600
00060700
00060800
00060900
00061000
00061100
00061200
00061300
00061400
00061500
00061600
00061700
00061800
00061900

FORTRAN IV C1	RELEASE 2.0	ANALYS	DATE - 80242	12/29/34	PAGE 0002
0044	NOPTC = NOPTC+MCT				00062300
0045	NOPTS = NOPTS+MCP				00062400
0046	CAA(1) = D(NOPTC)+AM(N1)				
0047	CBK(1) = D(NOPTC)+BK(N1)				
0048	SAA(1) = D(NOPTS)+AM(N1)				
0049	SBK(1) = D(NOPTS)+BK(N1)				
0050	166 CONTINUE				00061600
0051	NO 20	N1 = N1, N1			00061700
0052	CAL FREAD(X1), 30, N1B, 9999, 9999)				00061900
0053	NUC = (N1-1)*N1/2+N1				
0054	NO 20	N1 = 1, N1			
0055	CA = CAA(1)				
0056	CB = CBK(1)				
0057	SA = SAA(1)				
0058	SB = SBK(1)				
0059	NR = 1				
0060	17(N.L.E.NC2) GO TO 16				
0061	SA = -SA				
0062	SB = -SB				
0063	NR(NR, NUC) = NR(NR, NUC)+X(NR)*(CA+SB)				00062800
0064	NR(NR, NUC) = NR(NR, NUC)+X(NR)*(SA+CB)				00062900
0065	20 CONTINUE				00063000
0066	999 CONTINUE				00063100
0067	RETURN				
0068	END				

PAGE 0001

12/29/34

DATE - 00242

FWRITE

RELEASE 2.0

PORTMAN IV C1

0001 SUBROUTINE FWRITE(NOV, IO, R, S, *)
0002 LOGICALS:1 NOV
0003 DIMENSION NOV(N)
0004 WRITE(10) NOV
0005 RETURN
0006 END

PAGE 0001

12/29/34

DATE - 00242

FREAD

RELEASE 2.0

PORTMAN IV C1

0001 SUBROUTINE FREAD(NOV, IO, R, S, *)
0002 LOGICALS:1 NOV
0003 DIMENSION NOV(N)
0004 READ(10) NOV
0005 RETURN
0006 END

PAGE 0001

12/29/34

DATE - 00242

REWIND

RELEASE 2.0

PORTMAN IV C1

0001 SUBROUTINE REWIND(10)
0002 REWIND 10
0003 RETURN
0004 END

increasing m , stored in vector array form column by column; 30 stores the χ_i^{nm} of the optimal quadratures-type estimator. The χ_i^{nm} are stored from N. Pole to S. Pole, and according to nm , as the Legendre integrals and the coefficients. In some circumstances the grid may be geocentric rather than geodetic and a change of coordinates might be desirable: this can be achieved by setting the parameter IGEO to 2. The flattening assumed for this transformation is $F = 1/298.257$.

After the $R(m)$ matrices have been created, they are inverted by IMSL subroutine LUDECP, that performs a Choleskii factorization. IMSL subroutine LUELMP solves the equations resulting in the χ_i^{nm} ; if during the inversion LUDECP detects an ill-conditioned (or a singular) matrix, the solution part is avoided, and a set of null χ_i^{nm} is stored for that particular m . As an additional check for the stability of the solution, the relative residuals,

$$r = \frac{\sum_{i=0}^{N-1} v_i^2}{\sum_{i=0}^{N-1} (k_i^{nm})^2} \quad \text{where } \underline{v} = [v_0, v_1, \dots, v_{N-1}]^T \text{ is} \quad (B3)$$

$$\underline{v} = \underline{k}^{nm} - R(m) \underline{\chi}^{nm} \text{ (computed)}$$

are computed and printed. In all the cases studied here these residuals indicated an agreement of at least 9 significant figures. To improve the stability of the solution, a regularizing constant REGUL is added to the diagonal elements of the $R(m)$ (Paragraph (3.3)).

Arrays PN, SS, and FC contain the propagated noise, sampling, and total error measure (variance) per degree. W contains the averaged row variances (expression (2.43)) arranged from North to South.

The scalar arguments, NMAX, NN, DGRID, IGEO, REGUL, NRUN, and NC 2, are described in the comments inserted between statements 0006 and 0010. The arrays are as follows:

NAME	DIMENSION	TYPE
ROWP	$\frac{1}{2}(NMAX+1)(NMAX+2)$	REAL * 8
ROWQ	"	"
RHS	$\frac{1}{4}NN(NN+1)N$ ($N = \text{Nyquist freq.}$)	"
S	$(NN+1)N$	"
A	$\frac{1}{2}N(N-1)+N$	"
UL	"	"
W	N	"
DVAR	NMAX + 1	"
FC	NN + 1	"
PN	"	"
SS	"	"

Arrays ANMPQ, FF, XO, B, X, BT, FINMP (all REAL * 8), and IDD (REAL * 4),

all dimensioned 200 or 400 in the subroutine itself, are large enough for problems where $N < 200$. For finer grids, the size of these arrays should be increased in the same proportion as that of N .

The $R(m)$ matrices are formed according to expression (2.63). Subroutine FUR computes the "aliased" Fourier coefficients of the covariance functions that are, in fact, the elements of the $R(m)$, scaled by N or $2N$, depending on m . Common MM and array MT are part of a logic set up to ensure that the Fourier coefficients are not computed more than once each.

Subroutine ANALYS uses the x_i^{nn} stored in unit 30 to analyze the data in array DATA. The α_n^i, β_n^i are formed in place, as in HARMIN, so the original values in DATA are destroyed. Arrays CR, SR, CAA, CBB, SAA, and SBB have all the same description as CR1, SR1, etc., in HARMIN. IMSL subroutine FFCSIN (double precision) is used to obtain the α_n^i, β_n^i . The reason why ANALYS is used instead of HARMIN, is the arrangement of the x_i^{nn} "columnwise", or by increasing latitudes, rather than "rowwise" (i.e., all the x_i^{nn} for the same i stored together) as HARMIN would require.

The listings of FUR, ANALYS, and those of the fast input/output subroutines FREAD, FWRITE, REWIND, are given after that of NORMAL. The input/output subroutines have dummy arguments, because originally NORMAL was written to work with certain subroutines available at O.S.U. that may not be in the software libraries of other institutions.

B.5 Subroutine NORMAX

A modified version of NORMAL, this subroutine was created to compute the variance of the estimation errors in ordinary quadratures formulas according to the theory in section 2. Essentially, it computes

$$\sigma_{\epsilon_n}^2 = \sigma_n^2 - 2 \sum_{n=0}^N \sum_{\alpha=0}^I \underline{c}_n^T = \alpha_n : \underline{f}_{nn}^\alpha + \sum_{n=0}^N \sum_{\alpha=0}^I (\underline{f}_{nn}^\alpha)^T (C_{zz} + D) \underline{f}_{nn}^\alpha$$

by forming and using the $R(m)$. Since no inversion or solution of the normal equations is required, the corresponding segments have been removed from NORMAL, and a new final segment added for the computation of the various accuracies.

The theory behind the calculation of the $\sigma_{\epsilon_n}^2$ using the $R(m)$ matrices is as follows:

In the case of ordinary formulas of the type (see expression (1.7))

FORTRAN IV C1	RELEASE 2.0	NORMAX	DATE = 80242	12/29/84	PAGE 0000
0003	RC21 = RC2+1				00023000
0004	RC28 = 0*RC2				00023100
0005	FF(1) = DCRC+2				00023200
0006	FF2 = RI/2				
0007	BLK = DCRI+DCRCV				
0008	BLK = DSQRT(DSIN(BLK+BLK/PI)/DCRCV				
0009	CALL QUADPS(QUADS,RI1,BLCK,RC2)				
0010	BVANO = 0.00				
0011	DO 5 I = 1,RC				00023300
0012	FCI(1) = 0.00				00023400
0013	PR(1) = 0.00				
0014	5 CONTINUE				
0015					
0016					
0017					
0018					
0019					
0020					
0021					
0022					
0023					
0024					
0025					
0026					
0027					
0028					
0029					
0030					
0031					
0032					
0033					
0034					
0035					
0036					
0037					
0038					
0039					
0040					
0041					
0042					
0043					
0044					
0045					
0046					
0047					
0048					
0049					
0050					
0051					
0052					
0053					
0054					
0055					
0056					
0057					
0058					
0059					
0060					
0061					
0062					
0063					
0064					
0065					
0066					
0067					
0068					
0069					
0070					
0071					
0072					
0073					
0074					
0075					
0076					
0077					
0078					
0079					
0080					
0081					
0082					
0083					
0084					
0085					
0086					
0087					
0088					
0089					
0090					
0091					
0092					
0093					
0					

```

0072 358 FORMAT('' BLOCK SIZE = P7.2, MAX. DEGREE IN COVARIANCE = 15,
3 MAX. DEGREE TO BE ANALYZED = 15, NUMBER OF ROWS = 15, NRUB
3 15,')
0073 556 FORMAT(4,536) NC
0074 1F1(CEO,EO,1) WRITE(6,517)
0075 1F1(CEO,EO,2) WRITE(6,550)
0076 557 FORMAT(' DATA ON A SPHERE')
0077 558 FORMAT(' DATA ON A ELLIPSOID')
0078 DO 99 NC = 1, NPF1
0079 99 100 N1 = ((N1-1)*M1)/2
0080 DO 100 N1 = 1, 2*N1
0081 100 P7(N1) = 2.0*((1.0-DCOS((N1-1)*DCR))/(N1-1))**2
0082 READ THE R.N.S.'S OF THE REDUCED "NORMALS" FROM UNIT 10
INTO VECTOR RUS
C
C
C
0083 CALL REVIND(10)
0084 CALL REVIND(30)
0085 ISN = -NLL
0086 DO 210 N10 I = 1, NCA
0087 CALL FREAK,NOMP(1),10,NLLB,5999,5999)
0088 CALL FREAK,NOMX(1),30,NLLB,5999,5999)
0089 ISN = ISN+NLL
0090 DO 210 N10 J = 1, NLL
0091 XNS7(ISH,J) = NOMX(J)
0092 NNS(ISH,J) = NOMP(J)
0093 IF(NRUB,EO,99) GO TO 1000
C
C
C
0094 MAIN LOOP FOR THE CREATION OF THE R.N.D.'S MATRICES.
C
C
0095 CALL REVIND(10)
0096 CALL REVIND(30)
0097 CONTINUE
0098 DO 101 S(1) = 0.00 I = 1, ND
0099 101 S(1) = 0.00 POSITION THE INPP'S TAPE TO CURRENT ROW NP
C
C
C
0099 DO 102 RX = 1, NP
0100 CALL FREAK,NOMP(1),10,NPPO,5999,5999)
0101 IS = (NP-2)*NR1
0102 NCV = NCS-NP+1
0103 DO 120 NQ = NP, NCV
0104 IF(NQ,NE,NT) CALL FREAK,NOMX(1),10,NPPO,5999,5999)
0105 IF(NQ,NE,NT) GO TO 103
0106 DO 103 I = 1, NPP
0107 103 NOMX(1) = NOMP(1)
C
C
C
0108 OBTAIN THE FOURIER COEFFICIENTS OF THE BLOCK COVARIANCES
CORRESPONDING TO ROWS P AND Q IN CDD.
AND ALIASSE THEM TO FORM THE ELEMENTS OF THE R(N) MATRICES.
C
C
C
0108 DO 1 I = 1, 310
0109 UL(I) = 0.00
0110 HT(I) = 0
0111 I CONTINUE
C
C

```

```

0113 DO 111 N1 = 1,NN1
0114 N = N1-1
0115 N = NC2
0116 IF (N.EQ.0 OR N.EQ.NC2) N = NC
0117 CALL FUR(AN,N1,ROWP,ROWQ,IDD,FF,RP1)
0118 ANFG(N1) = AN
0119 N1 = NC+NP+1
0120 N2 = NC+NP+1
0121 CONTINUE
0122 IF (N1.GT.NP1) GO TO 111
0123 CALL FUR(AN,N1,ROWP,ROWQ,IDD,FF,RP1)
0124 ANFG(N1) = ANFG(N1)+AN1
0125 IF (N2.GT.NP1) GO TO 100
0126 CALL FUR(AN,N2,ROWP,ROWQ,IDD,FF,RP1)
0127 ANFG(N1) = ANFG(N1)+AN2
0128 N1 = N1+NC
0129 N2 = N2+NC
0130 GO TO 112
0131 ANFG(N1) = ANFG(N1)+E
0132 IS = IS+NN1
0133 DO 114 N1 = 1,NN1
0134 S(1S+NL) = ANFG(NL)
0135 NL = 1,NN1
0136 CONTINUE
0137 STORE THE LOWER DIAGONAL PART OF R(N) COLUMN BY COLUMN
ON DISK FILE (UNIT 15 ).
C
C
C
0137 DO 116 N2 = 1,NN1
0138 IS = -NN1
0139 DO 115 I = 1,NC2
0140 IS = IS+NN1
0141 RS(I) = S(1S+I2)
0142 CALL FWRITE(S(I),1S,NC2B,8999,8999)
0143 CONTINUE
0144 CALL REVINDX(6)
0145 NP = NP+1
0146 IF (NP.LE.NC4) GO TO 1000
0147 CONTINUE
C
C
0148 BEGIN MAIN LOOP WHERE ERRORS ARE ESTIMATED.
C
0148 DO 300 N1 = 1,NN1
0149 CALL REVINDX(5)
0150 PTF = PTF(N1)+NC
0151 N = N1-1
0152 IF (N.EQ.0) GO TO 214
0153 DO 213 I = 1,N
0154 CALL FREAD(S(I),1S,NC2B,8999,8999)
C
C
0155 READ IN MATRIX R(N) FROM UNIT 15 ; COLUMN BY COLUMN,
AND PUT IT IN SYMMETRIC STORAGE FORM IN VECTOR A .
C
0156 DO 220 N2=1, NC2-J+1
0157 IF (J.EQ.1) GO TO 222
0158 DO 215 I = 1,NN1
0159 CALL FREAD(S(I),1S,NC2B,8999,8999)

```


PORTMAN IV G1 RELEASE 2.0
 0214
 0215
 0216
 0217

NORWAY

DATE - 00242

12/29/34

PAGE 0006

999
 CALL REVIEWED (30)
 CONTINUE
 RETURN
 END

0005:1500
 0005:1500
 0005:1500
 0005:1500

$$\begin{aligned}
\hat{C}_{nn}^{\alpha} &= (\underline{f}_{nn}^{\alpha})^T \underline{m} \\
&= \mu_n \sum_{i=0}^{N-1} \sum_{j=0}^{2N-1} \int_{\sigma_{ij}} \bar{Y}_{nn}^{\alpha}(\theta, \lambda) d\sigma \, m_{ij} \\
&= \sum_{i=0}^{N-1} \left[\chi_i^{nn} \sum_{j=0}^{2N-1} \begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix} \cos m j \Delta \lambda + \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \sin m j \Delta \lambda \right] m_{ij}
\end{aligned}$$

the estimator vector is a combination of a "sine" and a "cosine" vector of "frequency" m (terminology introduced in paragraph (2.10)). The product of such vector by $(C_{zz} + D)$ is, because of the structure of this matrix, another combination of a "sine" and a "cosine" vector of the same "frequency". From the properties of the normal matrix follows that

$$(C_{zz} + D) \underline{f}_{nn}^{\alpha} = \begin{Bmatrix} 2N & \text{if } m=0 \\ & \text{or } m=N \\ N & \text{if } m \neq 0 \end{Bmatrix} \left[\begin{array}{c} \dots \nu_i^{nn} \begin{Bmatrix} A(m) \\ B(m) \end{Bmatrix} \cos m j \Delta \lambda + \\ \begin{Bmatrix} B(m) \\ A(m) \end{Bmatrix} \sin m j \Delta \lambda \dots \end{array} \right]^T$$

where, calling

$$\underline{\nu}^{nn} = [\nu_0^{nn}, \nu_1^{nn}, \dots, \nu_{N-1}^{nn}]^T = R(m) [\chi_0^{nn}, \chi_1^{nn}, \dots, \chi_{N-1}^{nn}]^T$$

and

$$F(m) = \begin{cases} 2N \Delta \lambda^2 & \text{if } m=0 \text{ or } m=N \\ 4N \frac{(1 - \cos m \Delta \lambda)}{m^2} & \text{if } m \neq 0, m \neq N \end{cases}$$

is

$$(\underline{f}_{nn}^{\alpha})^T (C_{zz} + D) \underline{f}_{nn}^{\alpha} = F(m) \sum_{i=0}^{N-1} \chi_i^{nn} \nu_i^{nn} \quad (B4-2)$$

Regarding the scalar product $2 \underline{c}_{nn}^T \alpha_z \underline{f}_{nn}^{\alpha}$, it is easy to show that

$$2 \underline{c}_{nn}^T \alpha_z \underline{f}_{nn}^{\alpha} = 2 \frac{\sigma_n^2}{2n+1} \sum_{i=0}^{N-1} (\chi_i^{nn})^2 F(m) \quad (B4-b)$$

Expressions (B4-a) and (B4-b) are implemented by NORMAX to obtain the error variances. This subroutine also uses FUR. Subroutine QUADRS is also called, to obtain the de-smoothing factors. In the case listed here, this factor is $\mu_n = \frac{1}{4\pi\beta_n}$. Array QUADS has been added to the list of arguments, and it contains the μ_n after the call to QUADFS.

B.6 Subroutine LEGFDN

This subroutine computes the normalized Legendre functions and, if so desired, their derivatives at a given θ_1 . All values correspond to the same order M ; if more than one order at a time is needed, a DO loop, where the subroutine is called once for each order, can be set up. The subroutine is based on formulas (4.19 a-b) and (1.38 a-b). The use of this subroutine is explained by the comments inserted in the listing. The stability of the recursive formulas was tested by computing $\bar{P}_{nn}(\cos \theta)$ and $(d\bar{P}_{nn}/d\theta)(\cos \theta)$ for $m = 350$ and $350 \leq n \leq 400$, $2.5^\circ \leq \theta \leq 90^\circ$, at 5° intervals. Calculations were done first in double precision (8 byte words) and then repeated in extended precision

SUBROUTINE LEAFDM(N,THETA,MLEG,DLEG,NMX,IR,RLNR,IPLAC)
 THIS SUBROUTINE COMPUTES ALL NORMALIZED LEGENDRE FUNCTIONS
 IN "MLEG" AND THEIR DERIVATIVES IN "DLEG". ORDER IS ALWAYS
 N, AND COLATITUDE IS ALWAYS THETA (RADIANS). MAXIMUM DEGREE
 IS NMX. ALL CALCULATIONS IN DOUBLE PRECISION.
 IR MUST BE SET TO ZERO BEFORE THE FIRST CALL TO THIS SUB.
 THE DIMENSIONS OF ALWAYS MLEG, DLEG, AND RLNR MUST BE
 AT LEAST EQUAL TO NMX+1
 IF THIS SUBROUTINE IS TO BE USED TO COMPUTE FUNCTIONS
 AND THEIR DERIVATIVES FOR MORE THAN ONE ORDER N, THEN
 THE HIGHEST ORDER SHOULD BE COMPUTED IN THE FIRST CALL.
 THIS PROGRAM DOES NOT COMPUTE DERIVATIVES AT THE POLES.
 IF IPLAC = 1, ONLY THE LEGENDRE FUNCTIONS ARE
 COMPUTED.

PROGRAMMER : OSCAR L. CUNOMO, DEPT. OF GEODETIC SCIENCE,
 THE OHIO STATE UNIVERSITY, AUGUST 1980. *****

IMPLICIT REAL*8 (A-H,O-Z)
 DIMENSION MLEG(1),DLEG(1),RLNR(1)
 2 DRTS(800),DIRT(800)
 NMX1 = NMX+1
 NMX2P = 2*NMX+1
 N1 = N+1
 N2 = N+2
 N3 = N+3
 IF(1N.EQ.1) GO TO 10
 IR = 1
 DO 5 N = 1,NMX2P
 DRTS(N) = DSORT(N+1,D0)
 5 DIRTN = 1.D0/DRTS(N)
 10 COTHET = DCOS(THETA)
 SITHET = DSIN(THETA)
 SITH1 = 1.D0/SITHET

COMPUTE THE LEGENDRE FUNCTIONS .

RLNR(1) = 1.D0
 RLNR(2) = SITHET*DRTS(3)
 DO 15 N1 = 3,N1
 N2 = 2*N
 15 RLNR(N1) = DRTS(N2+1)*DIRTN2*SITHET*RLNR(N1-1)
 IF(N.EQ.1) GO TO 20
 IF(N.EQ.2) GO TO 16
 RLNR(2) = RLNR(2)
 RLNR(3) = DRTS(5)*COTHET*MLEG(2)
 GO TO 20
 16 MLEG(1) = 1.D0
 MLEG(2) = COTHET*DRTS(3)
 20 CONTINUE
 RLNR(N1) = RLNR(N1)
 MLEG(N2) = DRTS(N+2+1)*COTHET*MLEG(N1)
 DO 30 N1 = N3,NMX1
 N = N1-1

PORTMAN IV G1 RELEASE 2.0 LEGEND DATE = 80242 13/02/38 PAGE 0002
 0035 IF(N.EQ.0.AND.N.LT.2.OR.N.EQ.1.AND.N.LT.3) GO TO 30
 0036 N2 = 2*N
 0037 NLEG(N1) = DRTS(N2+1)*DIRTY(N+D)*DIRTY(N-N)*COTHET*
 2 NLEG(N1-1) - DRTS(N+1)*DIRTY(N-N-1)*DIRTY(N2-3)*RLEG(N1-2)
 0038 GO TO 30
 0039 30 CONTINUE
 0040 IF(FLAG.EQ.1) RETURN
 0041 IF(SITHEQ.EQ.0.D0) WRITE(6,99)
 0042 99 FORMAT('/// *** LEGEND DOES NOT COMPUTE DERIVATIVES AT THE POLES
 2 *****')
 0043 IF(SITHEQ.EQ.0.D0) RETURN
 C
 C
 C
 0044 COMPUTE ALL THE DERIVATIVES OF THE LEGENDRE FUNCTIONS.
 0045 RLNR(1) = 0.D0
 0046 RLNR = RLNR(2)
 0047 RLNR(2) = DRTS(3)*COTHET
 0048 DO 40 N = 3, N1
 0049 N2 = N1-1
 0050 RLNR1 = RLNR(N1)
 0051 RLNR(N1) = DRTS(N2+1)*DIRTY(N2)*COTHET*RLNR(N)+COTHET*RLNR1
 0052 RLNR = RLNR1
 0053 40 CONTINUE
 0054 DLEG(N1) = RLNR(N1)
 0055 DO 60 N1 = N2, NMX1
 0056 N2 = N1-1
 0057 DLEG(N1) = SITHEQ(N *RLNR(N1))*COTHET-DRTS(N-N)*DRTS(N+N)*
 2 DRTS(N2+1)*DIRTY(N2-1)*RLEG(N1)
 0058 60 CONTINUE
 0059 RETURN
 0060 END
 0061

(16 byte words). Both sets of results agreed with each other to better than 10 significant figures.

B. 7 Subroutine NVAR

This subroutine computes the degree variances of the gravity anomalies (point values), up to degree 100, according to the coefficients obtained by R. Rapp from a complete, equal angular set of mean $1^\circ \times 1^\circ$ anomalies, as mentioned in par. (3.1). Above $n = 100$, the subroutine uses a two-term model to calculate the $\sigma_n^2(\Delta g)$. The resulting degree variances are stored in array DVAR. The first element in DVAR is $\sigma_1^2(\Delta g)$.

PAGE 0001

12/29/34

DATE - 00242

RVAR

FORTRAN IV G1 RELEASE 2.0

SUBROUTINE RVAR(DVAR, RMAX, RVAR0)

THIS SUBROUTINE COMPUTES ALL DEGREE VARIANCES UP TO
DEGREE "RMAX", USING SOME GIVEN MODEL.

C
C
C

IMPLICIT REAL*8 (A-H,O-Z)

INTEGER DVAR(1)

RVAR0 = 0.00

DO 10 I = 1, RMAX

RVAR(I) = 7.6396818

RVAR(2) = 33.962963

RVAR(3) = 19.743439

RVAR(4) = 19.696193

RVAR(5) = 18.990314

RVAR(6) = 18.485112

RVAR(7) = 10.399362

RVAR(8) = 8.6398739

RVAR(9) = 9.1031027

RVAR(10) = 6.3490710

RVAR(11) = 2.8993199

RVAR(12) = 6.7286206

RVAR(13) = 3.2033850

RVAR(14) = 3.8198187

RVAR(15) = 4.9444726

RVAR(16) = 3.7168750

RVAR(17) = 4.1988269

RVAR(18) = 3.4360349

RVAR(19) = 2.2931356

RVAR(20) = 3.0083072

RVAR(21) = 3.5082009

RVAR(22) = 2.6197612

RVAR(23) = 2.6080078

RVAR(24) = 3.4761637

RVAR(25) = 2.2875367

RVAR(26) = 2.1490026

RVAR(27) = 2.5005299

RVAR(28) = 2.0560312

RVAR(29) = 2.9207703

RVAR(30) = 2.4778694

RVAR(31) = 3.4971836

RVAR(32) = 3.2040108

RVAR(33) = 3.9118487

RVAR(34) = 8.1070939

RVAR(35) = 2.9400243

RVAR(36) = 3.0126138

RVAR(37) = 2.1799143

RVAR(38) = 2.8229174

RVAR(39) = 3.0049848

RVAR(40) = 2.8244301

RVAR(41) = 2.2442772

RVAR(42) = 3.0705790

RVAR(43) = 2.7283179

RVAR(44) = 3.9169446

RVAR(45) = 3.4346006

RVAR(46) = 2.8100541

RVAR(47) = 2.2040232

RVAR(48) = 3.4272534

RVAR(49) =

RVAR(50) =

00051000
00051000
00052000
00052100
00052200
00052300
00052400
00052500
00052600
00052700
00052800
00052900
00053000
00053100
00053200
00053300
00053400
00053500
00053600
00053700
00053800
00053900
00054000
00054100
00054200
00054300
00054400
00054500
00054600
00054700
00054800
00054900
00055000
00055100
00055200
00055300
00055400
00055500
00055600
00055700
00055800
00055900
00056000
00056100
00056200
00056300
00056400
00056500
00056600
00056700
00056800
00056900
00057000
00057100
00057200
00057300
00057400
00057500

PAGE 0002

12/29/84

DATE = 80242

POSTRAM IV G1 RELEASE 2.0

FVAR

0008 811 = 2.8176026
0009 812 = 2.6570577
0010 813 = 3.4827293
0011 814 = 3.0830738
0012 815 = 3.1636091
0013 816 = 2.8039749
0014 817 = 3.2189365
0015 818 = 2.8070499
0016 819 = 2.9280364
0017 820 = 3.2459338
0018 821 = 3.5989317
0019 822 = 3.3193559
0020 823 = 2.8970374
0021 824 = 2.7474396
0022 825 = 3.4163538
0023 826 = 2.8065284
0024 827 = 2.6995283
0025 828 = 2.8628832
0026 829 = 3.1792570
0027 830 = 2.4727662
0028 831 = 2.9711426
0029 832 = 2.7505264
0030 833 = 2.8784988
0031 834 = 2.9564934
0032 835 = 2.8971508
0033 836 = 2.8443190
0034 837 = 2.3198284
0035 838 = 2.4585835
0036 839 = 2.8181849
0037 840 = 2.8002385
0038 841 = 2.8064117
0039 842 = 3.1918590
0040 843 = 2.7445132
0041 844 = 3.0208313
0042 845 = 2.9415316
0043 846 = 2.4780426
0044 847 = 2.8778623
0045 848 = 2.8175941
0046 849 = 2.4938926
0047 850 = 2.4438377
0048 851 = 2.7012343
0049 852 = 2.8148959
0050 853 = 2.8470976
0051 854 = 2.8721000
0052 855 = 2.2715440
0053 856 = 2.8304260
0054 857 = 2.3360790
0055 858 = 2.6030127
0056 859 = 2.2631029
0057 860 = 2.3381125
0058 861 = IFIMAX.L.E.100) GO TO 20
0059 862 = A = 1.00
0060 863 = B = 2.00
0061 864 = A1 = 3.4050D0
0062 865 = A2 = 140.03D0
0063 866 = S1 = 0.998006D0
0064 867 = S2 = 0.914232D0
0065 868 = DO 10 N = 101, NMAX
0066 869 =
0067 870 =
0068 871 =
0069 872 =
0070 873 =
0071 874 =
0072 875 =
0073 876 =
0074 877 =
0075 878 =
0076 879 =
0077 880 =
0078 881 =
0079 882 =
0080 883 =
0081 884 =
0082 885 =
0083 886 =
0084 887 =
0085 888 =
0086 889 =
0087 890 =
0088 891 =
0089 892 =
0090 893 =
0091 894 =
0092 895 =
0093 896 =
0094 897 =
0095 898 =
0096 899 =
0097 900 =
0098 901 =
0099 902 =
0100 903 =
0101 904 =
0102 905 =
0103 906 =
0104 907 =
0105 908 =
0106 909 =
0107 910 =
0108 911 =
0109 912 =

0113
 0114
 0115
 0116
 0117

FORTRAN IV C1 RELEASE 2.0

EVAL

DATE = 00242

12/29/34

PAGE 0003

00063400
 00063500
 00063600
 00063700
 00063800

EVAL(N) = (N-1)*(S1+(N+2)*A1/(N+A)+S2*(N+2)*A2/(N+B)*(N-2)))
 10 CONTINUE
 20 CONTINUE
 RETURN
 END

B. 8 SUBROUTINE COVBLK

Subroutine COVBLK calculates covariances between area means according to expression (4.14). This subroutine is listed in the following pages. The arguments are explained in the comments at the beginning of the listing. In addition, the following things have to be born in mind: the dimension of the array DVAR is N_{\max} ; the dimension of both RINS and RINN is $2(N_{\max}+1)(N_{\max}+2)$; the dimension of COVS is $360/\text{BLOCK}$. The values returned in DVAR are the original degree variances c_n^{uv} , each divided by $2n+1$. If LB is less than $360/\text{BLOCK}$, the LB+1, LB+2, ..., $(360/\text{BLOCK} - \text{LB} - 1)$ elements in COVS are returned as zeroes, the remainder contain the first (and last) LB covariances. The dimension of F is $180/\text{BLOCK}$ (Nyquist frequency). To use this subroutine with $N_{\max} > 400$, the arrays FF and IDD (whose dimension should be no less than N_{\max}), should be redimensioned.

The subroutine does not take advantage of the "aliasing" of Fourier coefficients built into expression (4.14). Implementing this aspect should lead to some additional improvement in efficiency. The Fourier series is computed, once the coefficients have been determined, by multiplying each coefficient by the corresponding cosine of $m\lambda_j$ and adding the products together. The values of $\cos m\lambda_j$ are computed using the following recursive formula:

$$\cos m\lambda_j = 2 \cos \lambda_j \cos (m-1)\lambda_j - \cos (m-2)\lambda_j$$

which avoids repeated calculation of the FORTRAN COS function (only $\cos \lambda_j$ is required to start the recursion). Actual calculation of the Fourier series requires about 0.04 seconds in the most time consuming case: the grid of 1° blocks. The greater part of the time taken by this subroutine goes into finding the Fourier coefficients of the mean value covariances. For this reason, there is not much difference between computing all covariances in a certain row, or just a few of them, using this procedure.

FORTRAN IV C1 RELEASE 2.0 SUBROUTINE COVBK(DVAR, RINS, RINN, RMX, BLOCK, DLATS, DLATN, COVS, F, K, 19/02/38) DATE = 00243

THIS SUBROUTINE COMPUTES COVARIANCES BETWEEN AN AREA MEAN IN ROW "N" AND LB CONTIGUOUS AREA MEANS IN ROW "S" BEGINNING AT THE BLOCK WITH THE SAME LONGITUDE LIMITS AS THE ONE IN ROW "N".

DVAR : VECTOR OF DEGREE VARIANCES (PRECOMPUTED)
RINS : VECTOR OF INTEGRAL OF LEGENDRE FUNCTIONS, S. ROW (PRECOMPUTED)
RINN : VECTOR OF INTEGRALS OF LEGENDRE FUNCTIONS, N. ROW (PRECOMPUTED)
RMX : MAXIMUM DEGREE CONSIDERED IN THE EXPANSION OF THE COVARIANCE FUNCTION.
BLOCK : SIZE OF BLOCKS IN THE GRID (DEGREES)
LB : NUMBER OF CONTIGUOUS BLOCKS IN ROW "S".
DLATS : SOUTHWEST LAT. IN ROW "S" (SOUTH) (DEGREES)
DLATN : SOUTHWEST LAT. IN ROW "N" (NORTH) (DEGREES)
COVS : VECTOR OF COMPUTED COVARIANCES BETWEEN BLOCK MEANS
F : VECTOR OF FOURIER COSINE COEFFICIENTS OF THE TRIGONOMETRIC EXPANSION OF THE BLOCK MEANS.
K : COVARIANCE FUNCTION
INTEGER*4 CONSTANT SET TO 0 BEFORE FIRST CALL TO THIS SUBROUTINE.

PROGRAMMED BY OSCAR L. COLOMBO, DEPT. OF GEODETIC SC., OHIO STATE UNIVERSITY, COLUMBUS, DECEMBER 1979. *****

IMPLICIT REAL*8(A-H,O-Z)
DIMENSION DVAR(1), RINS(1), RINN(1), COVS(1), F(1), FF(400), IDDI(400), C(2,400)
IF(K.EQ.0) GO TO 10
RPI = RMX+1
F1 = 3.141592653589793D0
DRCOV = F1/100.D0
NB = 360.D0/BLOCK + 0.0000001D0
RBR = NB/2
RBRP = RBR+1
BLK = BLOCK*DRCOV
DO 1 I = 1, 2, RPI
1 FF(I) = 2.D0*(1.D0-DCOS((I-1)*BLK))/(I-1)**2
DO 2 I = 1, RPI
2 DVAR(I) = DVAR(I)/(2.D0*I-1.D0)
FF(I) = BLK**2
DO 3 I = 1, RPI
3 IDDI(I) = ((R1-I)*R1)/2
DO 4 I = 1, NB
4 C(I) = DCOS(BLK*(I-1))
K = 1
10 R1 = DLATS*DRCOV
R2 = DLATN*DRCOV
R1N = R1+BLK
R2N = R2+BLK
F1 = 1.D0/((DSIN(R1LN)-DSIN(R1N)) *BLK)*((DSIN(R2N)-DSIN(R2S))
2) *BLK)

COMPUTE DE COSINE FOURIER COEFFICIENTS OF THE BLOCK

FORTRAN IV C: RELEASE 2.0 COVBLEK DATE = 80242 13/02/38 PAGE 0002
 C COVARIANCE FUNCTION
 0027 RA = 0.00
 0028 DO 20 NI = 1, NP1
 0029 F(NI) = 0.00
 0030 RXX = NI
 0031 DO 10 NI = NI, NP1
 0032 IX = IDEX(NI)*NI
 0033 F(NI) = F(NI)*NINX(IX)*DVAR(NI)
 0034 F(NI) = F(NI)*FF(NI)*FR
 0035 FZ = F(NI)*2
 0036 NA = RAXZ
 0037 RAG = RAXZ
 0038 IF(NI.GT.1.AND.F(NI-1)*2.LT.RAG.AND.FZ.LT.RAG) GO TO 21
 0039 CONTINUE
 C
 C FIND THE COVARIANCES BETWEEN BLANKS BY FOURIER SYNTHESIS.
 0040 DO 40 I = 1, NBLP
 0041 COVS(I) = 0.00
 0042 IF(I.GT.10) GO TO 40
 0043 GRN1 = C(I)
 0044 C2 = 2.00/C(I)
 0045 C3 = 1.00
 0046 COVS(I) = F(I)
 0047 DO 30 NI = 2, NIX
 0048 GRN1 = C2*GRN1-GRN1
 0049 C3 = C3
 0050 C3 = C3*NI
 0051 COVS(NI) = COVS(I)*C3*FF(NI)
 0052 CONTINUE
 0053 DO 45 I = 1, NIB
 0054 COVS(NB-I+1) = COVS(I+1)
 0055 CONTINUE
 0056 RETURN
 0057 END